Entwicklung und Validierung eines Tests zur Messung des divergenten Denkens von Schülern

Ein Beitrag zur empirischen Kreativitätsforschung

Inaugural-Dissertation
zur Erlangung des akademischen Grades eines
Doktors der Philosophie
der Philosophischen Fakultät
der Universität des Saarlandes

vorgelegt von
Ursula Mainberger
aus Karlsruhe

1977

Entwicklung und Validierung eines Tests zur Messung des divergenten Denkens von Schülern

Ein Beitrag zur empirischen Kreativitätsforschung

Inaugural-Dissertation
zur Erlangung des akademischen Grades eines
Doktors der Philosophie
der Philosophischen Fakultät
der Universität des Saarlandes

vorgelegt von
Ursula Mainberger
aus Karlsruhe

UNIV.
LEIDEN
BIBL.

Der Dekan: Prof. Dr. L. Kötter

Berichterstatter: Prof. Dr. I. Strauch
Prof. Dr. K. Ingenkamp
Prof. Dr. W. H. Tack

Tag der letzten Prüfung: 2o. 12. 1977

Vorwort

Die vorliegende Arbeit entstand auf Anregung von Prof. Dr. K. Ingenkamp, dem ich für die mir während des gesamten Untersuchungszeitraums gewährte wissenschaftliche Beratung und Unterstützung danken möchte.

Ebenso bedanke ich mich bei Prof. Dr. I. Strauch für die Bereitschaft, diese Arbeit zu betreuen, für die wissenschaftliche Beratung und emotionale Unterstützung während der gesamten Bearbeitungszeit.

Allen Mitarbeitern des Zentrums für empirische pädagogische Forschung der Erziehungswissenschaftlichen Hochschule Rheinland-Pfalz schulde ich Dank für das Interesse, das sie meiner Arbeit entgegenbrachten, und für vielfältige Hilfestellungen bei der Durchführung der Untersuchungen.

Schließlich gilt mein Dank allen an den vielen Untersuchungen Beteiligten, vor allem den Schülern, Lehrern und den Studenten der Erziehungswissenschaftlichen Hochschule, deren Einsatzbereitschaft die Durchführung der mit großem Arbeitsaufwand verbundenen Untersuchungen ermöglichte.

Inhalt

		Seite
Vorwort		I
Literaturteil		1
1	Vorbemerkung	1
2	Verfahren zur Messung der Kreativität von Schulkindern	3
2.1	Guilfords theoretischer Beitrag zur Kreativitätsmessung	4
2.11	Guilfords Modell der Struktur des Intellekts	4
2.12	Kritische Einwände gegen die Theorie Guilfords	6
2.13	Kritische Einwände gegen die empirische Überprüfung der Theorie	9
2.14	Kritische Einwände gegen Guilfords Aufgaben zur Messung kreativer Fähigkeiten	11
2.15	Schlußfolgerungen für die weitere Arbeit	15
2.2	Die Torrance Tests of Creative Thinking (TTCT)	15
2.21	Theoretische Fundierung	16
2.22	Testgütekriterien	24
2.221	Objektivität	24
2.222	Reliabilität	25
2.223	Validität	32
2.2231	Inhaltsvalidität	32
2.2232	Übereinstimmungs- und Vorhersagevalidität	33
2.2233	Konstruktvalidität	44
2.23	Ausgewählte Untersuchungsergebnisse	55
2.231	Auswirkungen unterschiedlicher Testbedingungen	55
2.232	Testauswertung	6o
2.233	Geschlechtsabhängigkeit der TTCT-Ergebnisse	65
2.234	Schichtabhängigkeit der TTCT-Ergebnisse	68
2.235	Zusammenhänge zwischen TTCT-Ergebnissen, Intelligenz und Schulleistungen	72
2.24	Schlußfolgerungen für die weitere Arbeit	76

Seite

2.3 Der Kreativitätstest von Wallach und Kogan..... 77

2.31 Theoretische Fundierung..... 77

2.32 Testgütekriterien..... 86

2.321 Objektivität..... 88

2.323 Validität..... 93

2.3231 Konstruktvalidität..... 93

2.3232 Übereinstimmungs- und Vorhersagevalidität.... 11o

2.33 Ausgewählte Untersuchungsergebnisse..... 118

2.331 Auswirkungen unterschiedlicher Testbedingungen 118

2.332 Testauswertung..... 129

2.333 Geschlechtsabhängigkeit der Testergebnisse.... 135

2.334 Schichtabhängigkeit der Testergebnisse..... 138

2.335 Zusammenhänge zwischen den Testergebnissen, Intelligenz und Schulleistung..... 14o

2.34 Schlußfolgerungen für die weitere Arbeit..... 141

Empirischer Teil..... 144

3 Entwicklung und Validierung eines Verfahrens zur Erfassung des divergenten Denkens für Schüler der BRD..... 144

3.1 Theoretische Fundierung..... 144

3.2 Testentwicklung..... 149

3.21 Erste und zweite Versuchsform (II, 4 und II, 5)..... 15o

3.22 Dritte Versuchsform (IV, 1)..... 153

3.221 Untersuchung an 15 Gesamtschulklassen..... 154

3.222 Untersuchungen an 4 Hauptschulklassen..... 164

3.2221 Ergebnisse der Untersuchung mit dem Lehrerfragebogen..... 165

3.2222 Ergebnisse der Schülerbefragung..... 178

3.2223 Ergebnisse der Untersuchung mit dem Kreativitätstest von Wallach und Kogan..... 187

3.223	Folgerungen aus den Untersuchungsergebnissen..	191
3.2231	Überlegungen zu den Validitätsuntersuchungen..	191
3.2232	Revision der Versuchsform IV, 1.....	193
3.3	Testvalidierung.....	196
3.31	Überblick über die Untersuchungen.....	196
3.32	Untersuchungsziele.....	197
3.321	Untersuchungen in der Gesamtstichprobe.....	198
3.3211	Überprüfung der diskriminanten Validität des Verfahrens.....	198
3.3212	Überprüfung der Geschlechtsabhängigkeit des Verfahrens.....	199
3.3213	Überprüfung der Schichtabhängigkeit des Verfahrens.....	199
3.3214	Überprüfung der diagnostischen Valenz der Originalitätswerte.....	2oo
3.322	Untersuchungen in der Teilstichprobe.....	2o1
3.3221	Überprüfung der Übereinstimmungsvalidität mit dem Lehrerfragebogen.....	2o1
3.3222	Überprüfung der Übereinstimmungsvalidität mit der Schülerbefragung.....	2o2
3.3223	Überprüfung der Konstruktvalidität mit dem Test von Wallach und Kogan und zwei Intelligenztests.....	2o2
3.3224	Überprüfung der Konstruktvalidität mit dem Persönlichkeitsfragebogen CPQ.....	2o3
3.33	Darstellung und Diskussion der Ergebnisse.....	213
3.331	Ergebnisse der Gesamtstichprobe.....	213
3.3311	Diskriminante Validität des Verfahrens.....	213
3.3312	Geschlechtsabhängigkeit der Testergebnisse....	221
3.3313	Schichtabhängigkeit der Testergebnisse.....	223

Seite

3.3314 Diagnostische Valenz der Originalitätswerte... 226

3.332 Ergebnisse der Teilstichprobe..... 228

3.3321 Stichprobenmerkmale..... 228

3.3322 Ergebnisse der Untersuchung mit dem Lehrerfragebogen..... 231

3.3323 Ergebnisse der Untersuchung mit den Schülerfragen..... 237

3.3324 Ergebnisse der Untersuchung mit dem Test von Wallach und Kogan und 2 Intelligenztests..... 24o

3.3325 Ergebnisse der Untersuchung mit dem Persönlichkeitsfragebogen CPQ..... 244

4 Zusammenfassung der Arbeit und Wertung der Ergebnisse..... 251

Literaturverzeichnis..... 255

Anhang A

Anhang B

Literaturteil

1 Vorbemerkung

Obwohl der Begriff "Kreativität" im letzten Jahrzehnt in der deutschsprachigen psychologischen und erziehungswissenschaftlichen Forschung ständig mehr Beachtung fand[1] und obwohl bereits konkrete Maßnahmen zur schulischen Kreativitätsförderung vorgeschlagen wurden[2], deutet bisher nichts darauf hin, daß Überlegungen zu ihrer Realisierung in die Bildungsplanung aufgenommen würden.
Dafür gibt es sicher mehrere Gründe: Abgesehen von der derzeitigen allgemeinen Reformmüdigkeit im Bildungsbereich scheinen die Aktivitäten der extremen Linken zu einer verstärkten Wertschätzung angepaßten konformen Verhaltens geführt zu haben, die eine denkbar ungünstige Voraussetzung für Innovationen mit dem Ziel der Kreativitätsförderung darstellt.

Ein weiterer Grund dafür, daß ein weithin für wichtig gehaltenes Erziehungsziel in der Bildungspolitik keine Berücksichtigung findet, dürfte der Mangel an validen und ökonomischen Verfahren zur Überprüfung der Effektivität kreativitätsfördernder Unterrichtsmethoden sein. Daß in der BRD keine solchen Meßinstrumente zur Verfügung stehen, stellt gerade angesichts der gegenwärtigen Tendenz,

---

1) 1966 Flechsig
1968 Ulmann
1969 Grote, Hajek u.a., Kemmler, Landau, Lückert, Massialas und Zevin, Mutschler
1970 Brühl, Kerbs, Viebahn
1971 Gross, Schreiner, Wollschläger
1972 Begemann, Brickenkamp, Krause
1973 Brickenkamp, Herbig und Hollensteiner, Schiffler, Ulmann
1974 Fiedler und Windheuser, Heinelt, Hollensteiner, Laux und Zinn, Matussek, Rahm, Seiffge-Krenke, Seisenberger, Stammeier, Ulmann

2) z.B. Begemann 1972, Brühl 1970, Flechsig 1966, Herbig und Hollensteiner 1973, Hollensteiner 1974, Kemmler 1969, Landau 1969, Lückert 1969, Mainberger 1975, Massialas und Zevin 1969, Mühle und Schell (Hrsg.) 1970, Mutschler 1969, Schreiner 1971, Stammeier 1974, Viebahn 1970, Williams 1971 (diese unvollständige Aufzählung enthält auch einige aus dem Amerikanischen übersetzte Publikationen.)

nur operationalisierte oder zumindest leicht operationalisierbare Lernziele zu akzeptieren, eine äußerst ungünstige Bedingung für die Aufnahme des Lernziels "Kreativität" in die Curriculumplanung dar.
Eine wichtige Voraussetzung, bessere Bedingungen für schulische Kreativitätsförderung zu schaffen, besteht also darin, ein Verfahren zur Kreativitätsmessung bei Schulkindern zu entwickeln.

Wenn man eine Umgewichtung der derzeitigen Erziehungsziele zugunsten einer Förderung der kreativen Fähigkeiten der Kinder erreichen möchte, kann die Entwicklung eines Kreativitätstests allerdings nur als erster Schritt auf dem Weg zu diesem Ziel betrachtet werden. Wesentlich wichtiger dürfte der Nachweis eines Zusammenhangs zwischen der Kreativität der Schüler und späteren Leistungskriterien sein, denn in der heutigen, durch die Auswirkungen der Hochschulzulassungsbeschränkung und der Unsicherheit auf dem Arbeitsmarkt gekennzeichneten Bildungssituation dürften nur Lernziele, deren Bedeutung zumindest für den Schulerfolg nachgewiesen wurde, Aussicht auf Berücksichtigung haben[1)]. Dieser Nachweis der prädiktiven Validität eines Verfahrens kann aber nur mit Hilfe aufwendiger Längsschnittuntersuchungen erbracht werden und erfolgt daher aus forschungsökonomischen Gründen meist erst dann, wenn die Überprüfung der Übereinstimmungsvalidität zu befriedigenden Ergebnissen geführt hat.

Der dargestellten Zielsetzung entsprechend wird in der vorliegenden Arbeit zunächst untersucht, ob es sinnvoll ist, einen für Schüler geeigneten amerikanischen Kreativitätstests für bundesdeutsche Kinder zu adaptieren. Falls diese Frage aufgrund einer eingehenden Literaturanalyse verneint werden muß, soll versucht werden, ein Verfahren zur Erfassung kreativer Fähigkeiten für deutsche Schüler zu entwickeln und zu validieren.
In beiden Fällen würden sich die Untersuchungen auf eine Überprüfung der Konstrukt- und der Übereinstimmungsvalidität beschränken und können daher nur als ein Teil der erforderlichen Validierung betrachtet werden; der wichtige Nachweis der prädiktiven Validität kann im Rahmen dieser Arbeit nicht erbracht werden.

---

1) Zu anderen wichtigen Zielen einer schulischen Kreativitätsförderung s. Mainberger 1975, S. 232 - 233

Somit stellt die Gesamtheit der durchgeführten Untersuchungen im günstigsten Fall einen kleinen Schritt auf dem Weg zu dem Ziel dar, dem Lernziel Kreativität in unseren Schulen zu einer stärkeren Berücksichtigung zu verhelfen.

2 Verfahren zur Messung der Kreativität von Schulkindern

In diesem Kapitel wird ein Überblick über die für 9- bis 12jährige Schüler geeigneten Verfahren zur Erfassung kognitiver Voraussetzungen für kreatives Verhalten gegeben[1].

Eine Beschränkung auf die für jüngere Schüler entwickelten Tests erscheint deshalb angezeigt, weil hier der Validitätsnachweis auf andere Weise erbracht werden muß als bei Erwachsenen und älteren Jugendlichen, bei denen eventuell die Möglichkeit besteht, reale kreative Leistungen als Außenkriterien zu verwenden.

Da dieser Literaturüberblick unter Umständen Grundlage und Ausgangspunkt für die Konstruktion eines eigenen Verfahrens werden soll, wird zu jedem Verfahren ein möglichst breites Spektrum an Untersuchungsergebnissen zu theoretischen und zu pragmatischen Fragen der Kreativitätsmessung dargestellt. Unter diesem Blickwinkel erscheint auch die Erwähnung von Untersuchungsergebnissen mit negativem Resultat bei bestimmten Fragestellungen angezeigt, weil dadurch eventuell die Beurteilung und Einordnung von Untersuchungsergebnissen zu dem neu entwickelten Verfahren erleichtert werden.

Entsprechend der Absicht, die Verfahren relativ eingehend auf ihre Verwendbarkeit auf der hier interessierenden Altersstufe zu überprüfen, werden nur Tests besprochen, mit denen bereits eine größere Zahl von empirischen Untersuchungen an jüngeren Schülern durchgeführt worden sind, so daß zumindest Anhaltspunkte für die Beantwortung der wichtigsten Fragen vorliegen.

Da die von Guilford entwickelten Aufgaben zur Erfassung kreativer Fähigkeiten[2] fast ausschließlich an älteren Schülern erprobt[3]

---

1) Einen Überblick über die gesamte amerikanische Kreativitätsforschung bis 1967 geben Landau (1971[2]) und Ulmann (1968); Seiffge-Krenke (1974) stellt auch neuere Ergebnisse - z.T. bis 1971 - dar.

2) Kaltsounis (1971) führt in seiner Zusammenstellung von "Verfahren zur Untersuchung kreativen Verhaltens und kreativer Fähigkeiten" 16 von Guilford und seinen Mitarbeitern entwickelte Testaufgaben an. Diese Aufgaben werden beschrieben in Guilford und Hoepfner (1971), Seiffge-Krenke (1974) und Ulmann (1968).

3) Zumindest 3 Untersuchungen sind auszunehmen: Guilford und Hoepfner (1971) berichten über eine Studie, in der 4oo Schüler sechster Klassen mit diesen Aufgaben getestet wurden, wobei nur 4 von 8 erwarteten Faktoren identifiziert wurden. Weniger umfassende Zusammenstellungen von Guilford - Aufgaben wurden von Bennett (1973) an Zehnjährigen und von Guilford und Christensen an Schülern vierter, fünfter und sechster Klassen erprobt.

oder in wechselnder Zusammenstellung und z.T. zusammen mit anderen Kreativitätstest-Aufgaben verwendet wurden[1], werden sie hier nicht behandelt.
Sein theoretischer Beitrag zur Messung kreativer Fähigkeiten soll dagegen wegen seiner grundlegenden Bedeutung für die Fragestellung der vorliegenden Arbeit besprochen werden.

2.1 Guilfords theoretischer Beitrag zur Kreativitätsmessung

2.11 Guilfords Modell der Struktur des Intellekts

Guilford unterscheidet in seinem Strukturmodell der Intelligenz 5 Arten von (Denk-)Operationen - Kognition, Gedächtnis, divergente Produktion, konvergente Produktion und Bewertung-, die 6 verschiedene Produkte - Einheiten, Klassen, Relationen, Systeme, Transformationen und Implikationen - hervorbringen können. Die 5 Operationen können sich in verschiedenen "Inhalts"-Klassen ("Materialklassen" wäre eventuell eine eindeutigere Bezeichnung) abspielen; Guilford unterscheidet 4 Inhalte, nämlich figurale, symbolische, semantische und Verhaltens-Inhalte ("behavioral" im Sinn von Sozialverhalten).
Dieses dreidimensionale Klassifikationssystem[2] umfaßt somit 12o postulierte spezielle Fähigkeiten des menschlichen Intellekts, die Guilford durch zu diesem Zweck entwickelte Testaufgaben mit Hilfe der Faktorenanalyse nachzuweisen versucht. Er betrachtet alle 12o Intelligenzfaktoren als voneinander unabhängig, weshalb er orthogonale Rotationen durchführt. In Guilfords Modell sind eine große Anzahl kognitiver Voraussetzungen für kreative Leistungen berücksichtigt; unter diesen nehmen die Fähigkeiten zur divergenten Produktion, die durch Aufgaben mit freier Antwortform gemessen werden, eine zentrale Stellung ein[3].

---

1) Z.B. Cline, Richards and Needham 1963, Cropley 1966, Hargreaves and Bolton 1972, Hasan and Butcher 1966, Klausmeier, Harris and Ethnatics 1962, Piers, Daniels and Quackenbush 196o, Stubbings 1967, Sullivan 1973.

2) Ob Guilford den Begriff "Modell" für dieses Klassifikationssystem zu Recht in Anspruch nimmt, soll hier nicht erörtert werden. Nach Herrmann und Stapf (197o) handelt es sich bestenfalls um ein Modell im weiteren Sinn, während es nach Mayntz (1967, zitiert nach Herrmann und Stapf 197o) den deskriptiven - im Gegensatz zu den erklärenden - Modellen zugeordnet werden müßte.

3) Die für kreative Leistungen bedeutsamen Faktoren werden z.B. von Guilford 197o, Guilford and Hoepfner 1971 und Ulmann 1968 beschrieben und sollen daher hier nicht im einzelnen dargestellt werden.

Dem Vorwurf, er betrachte die Fähigkeiten zu divergentem Denken als die alleinige Voraussetzung für kreatives Verhalten, kann Guilford (1971) mit dem Hinweis begegnen, daß er von Anfang an[1] auf eine Reihe weiterer Fähigkeiten hingewiesen habe, die ihren Beitrag beim Zustandekommen schöpferischer Produktivität leisten. So vertritt er z.B. 1967 die Ansicht, daß Transformationsfähigkeiten - auch solche außerhalb der Kategorie der divergenten Produktion - eine wichtige Rolle spielen, 197o verweist er in diesem Zusammenhang auf alle 2o Transformationsfaktoren (von denen nur 4 zur DP-Kategorie gehören).

Guilfords unbestreitbares Verdienst ist es, die empirische Kreativitätsforschung ausgelöst und durch sein Klassifikationssystem der Denktätigkeiten ein Modell von großem heuristischem Wert geschaffen zu haben. Seine Mitarbeiter und er haben inzwischen etwa 1oo der 12o hypothetischen Faktoren nachgewiesen, wobei es sich als notwendig erwies, einigen der "Zellen" des Modells mehr als einen Faktor zuzuordnen[2]. Die Vielzahl der in dem Modell enthaltenen "Kreativitätsfaktoren", deren Kombination je nach Problemstellung ganz verschieden sein kann, trägt der Komplexität der vielfältigen und unterschiedlich verlaufenden kreativen Prozesse viel besser Rechnung als alle Kreativitätstheorien, die das Zustandekommen kreativer Leistungen durch einen oder einige wenige Faktoren zu erklären versuchen. Guilfords Annahme kontinuierlich verteilter, unabhängig variierender Fähigkeitsdimensionen wird gestützt durch die relative Seltenheit der Individuen, die bedeutsame kreative Leistungen erbringen.

Für den Testkonstrukteur ist überdies eine Kreativitätstheorie, die verschiedene nichtverbale Fähigkeitsbereiche berücksichtigt, von besonderem Interesse, da angenommen werden kann, daß ein auf dieser

---

1) Z.B. schon in seinem 195o vor der American Psychological Association gehaltenen Vortrag (s. Mühle und Schell 1971[2], S. 13 - 36).

2) Guilford hatte schon 197o 98 der 12o S-I-Faktoren demonstriert und zwar alle 24 Kognitions-Faktoren, 18 der 24 Gedächtnis-Faktoren (hier fehlt die Verhaltens-Kategorie noch vollständig), immerhin 23 der 24 Faktoren der divergenten Produktion (bisher konnte nur der Faktor "divergente Produktion figuraler Relationen" nicht nachgewiesen werden), 15 der 24 Faktoren der konvergenten Produktion und 18 Bewertungsfaktoren. Der semantische Inhaltsbereich ist vollständig erfaßt, für die 3o Zellen der erst später in das Modell aufgenommenen Verhaltenskategorie lagen 197o noch keine Aufgaben für die Gedächtsnis-Operationen und für die konvergenten und Bewertungsoperationen vor (s. Guilford 197o).

Grundlage entwickeltes Verfahren einer größeren Zahl von Begabungsrichtungen gerecht wird und daß jüngere Schüler durch eine Reihe verschiedenartiger Aufgaben stärker motiviert werden als durch einen rein verbalen Test zur Untersuchung semantischer und/oder symbolischer Fähigkeiten. Außerdem scheint Guilfords Validierungsmethode zunächst weniger Schwierigkeiten zu bieten als beispielsweise die Schaffung von Außenkriterien für kreative Fähigkeiten jüngerer Schüler.
Trotzdem gibt es eine Reihe von Überlegungen, die dagegen sprechen, Guilford-Aufgaben für jüngere Altersstufen zu adaptieren oder ein Verfahren auf der Grundlage seiner Theorie und mit Hilfe seiner Validierungsmethode zu entwickeln. Sie sollen im folgenden besprochen werden.

2.12 Kritische Einwände gegen die Theorie Guilfords

Guilfords Modell der Struktur des Intellekts geht von einer großen Zahl sehr spezieller und voneinander unabhängiger kognitiver Fähigkeiten aus. Seine Theorie sagt nichts über die Entwicklung dieser verschiedenartigen Fähigkeiten aus und die nichthierarchische Struktur des Modells bietet keine Ansatzmöglichkeiten für Hypothesen über Entwicklungsprozesse.

Sowohl die 3 Dimensionen des Klassifikationssystems als auch die verschiedenen Klassen wurden nicht empirisch, sondern logisch-intuitiv gewonnen[1]. Die Kategorisierung der Tätigkeiten des menschlichen Intellekts ist somit nicht zwingend; es wären andere Dimensionen denkbar und je nach Abstraktionsniveau könnten pro Dimension mehr oder weniger Klassen postuliert werden[2]. Fraglich erscheint es auch, ob die Denkprodukte der später in das Modell aufgenommenen Inhaltskategorie Sozialverhalten durch die Kategorien "Einheiten", "Klassen", "Relationen","Systeme", "Transformationen " und "Implikationen" am besten beschrieben werden oder ob hier nicht andere Produktklassen angemessener wären.

---

1) S. Wewetzer 1972

2) S. dazu Butcher 1973, Cattell 1971 (zitiert nach Butcher 1973) und Eysenck 1967, der bereits 1953 ein "model of the structure of intellect" veröffentlichte, das in 2 seiner 3 Dimensionen - einer "mental-process"- und einer "test-material"-Dimension - dem S-I-Modell Guilfords sehr ähnlich ist; Eysenck hat diese beiden Dimensionen allerdings in eine geringere Zahl von Kategorien untergliedert als Guilford.

Ein anderer Einwand richtet sich gegen Guilfords Annahme der Unabhängigkeit der Faktoren. In vielen Untersuchungen wurden mäßige bis mittlere positive Korrelationen zwischen Guilford-Tests zur Messung verschiedener Faktoren gefunden, was von vielen Vertretern hierarchischer Intelligenz-Theorien und Anhängern eines g-Faktors als Bestätigung ihres Standpunktes gewertet wurde[1]. Guilford (1964) versucht, die Kritik an seiner Theorie dadurch zu entkräften, daß er die Verteilung von über 7ooo Korrelationskoeffizienten aus 13 für seine Arbeit typischen Untersuchungen darstellt; etwa 24 Prozent der Koeffizienten sind - bei einer Irrtumswahrscheinlichkeit von .o5 - nicht signifikant von Null verschieden. Die Argumentation gegen die Annahme eines g-Faktors wird nach Darstellung dieser Zahlen nicht wieder aufgenommen, Guilford betont lediglich, daß bei einem so hohen Prozentsatz von Nullkorrelationen gute orthogonale Faktorenlösungen nach dem Prinzip der Einfachstruktur möglich sind. Insgesamt belegen Guilfords Angaben seine Annahme völlig unabhängiger Faktoren nicht allzu überzeugend; Eysenck ist der Meinung, daß Guilford durch die mangelnde Berücksichtigung der positiven Korrelationen zwischen den meisten Tests und zwischen verschiedenen Faktoren "has truly cut out the Dane from his production of Hamlet" (Eysenck 1967, S. 82).
Von einigen Kritikern Guilfords wurde die Beziehung zwischen Testaufgaben, wie sie in konventionellen Intelligenztests und bei Guilford in den Operations-Klassen Kognition, konvergente Produktion und Bewertung vorkommen, und Aufgaben zur Erfassung der divergenten Produktion (DP) untersucht. Da diese Beziehung als Spezialfall der Korrelationen aller S-I-Tests untereinander zu betrachten ist, wurden auch hier mäßige bis mittlere positive Korrelationen gefunden[2]. Guilford (1971) kann zwar Thorndikes Kritik an 2 seiner Untersuchungen zurückweisen, u.a. mit dem Hinweis, daß Thorndikes Methode der Gegenüberstellung der Korrelationen der Intelligenztestaufgaben untereinander, der DP-Tests untereinander und der Korrelationen der DP-Tests mit den Intelligenzmaßen ein angesichts der Komplexität der Situation zu einfaches Verfahren sei und verweist auf die in diesem Fall angemessene Methode der Faktorenanalyse. Aber Cropley (1966), der 3 Guilford-Aufgaben[3] und 2 Untertests von Torrance zusammen mit mehreren IQ-Maßen fak-

---

1) Z.B. Eysenck 1967, McNemar 1964, Vernon 1964, 1969

2) S.z.B. Cropley 1966, McNemar 1964, Thorndike 1962, Wallach 197o

3) Cropley entnahm diese Aufgaben dem "Manual for Kit of Reference Tests for Cognitive Factors" von French, Ekstrom and Price (1963), weswegen er als Testautor etwas irreführend "French 1963" angibt.

toranalysierte, erhielt eine im Sinn der Einfachstruktur bessere Faktorenlösung bei obliquer statt orthogonaler Rotation. Die Korrelation zwischen dem 1. Faktor, auf dem die Intelligenztests hoch luden, und dem 2., dem "divergenten" Faktor, betrug .514.

Auf die positive Beziehung zwischen DP-Testergebnissen und traditionellen Intelligenzmaßen weist Guilford selbst bereits 1966 hin. In einer Untersuchung, die er 1966 zusammen mit Hoepfner durchführte[1], waren die Streuungsdiagramme für viele Korrelationen zwischen DP-Tests und IQ dreieckig, d.h. Individuen mit niedrigem IQ hatten sehr selten hohe DP-Scores, während bei hochintelligenten Probanden die DP-Testergebnisse von niedrig bis hoch streuten[2]. Guilford und Christensen (1973) erklären dieses Ergebnis damit, daß Intelligenztests vor allem Kognitions-Operationen erfassen, die von den jeweils verfügbaren Informationen abhängen. Die Leistungen in DP-Tests hängen aber z.T. ebenfalls vom Vorhandensein aufgabenrelevanter Informationen ab, so daß Kognitionsfähigkeiten eines bestimmten Inhaltsbereichs eine obere Grenze für die divergente Produktion in demselben Bereich setzen. Die aus diesen Überlegungen abgeleitete Untersuchungshypothese, daß wegen des überwiegend semantischen Inhalts von Intelligenztests für die Beziehung zwischen IQ und semantischen DP-Tests höhere Korrelationen und dreieckige Streuungsdiagramme, für den Zusammenhang zwischen IQ und visuell-figuralen DP-Tests niedrigere (positive) Korrelationen und elliptische Streuungsdiagramme auftreten müßten, wurde durch die Ergebnisse im großen und ganzen bestätigt.

Interessant im Hinblick auf Guilfords Annahme unabhängiger Faktoren ist seine Erklärung der Dreiecksform der Streuungsdiagramme, die eine hierarchische Struktur der Denkoperationen impliziert, da divergente Produktionen ohne die entsprechenden Kognitionen nicht möglich sind. Diese Überlegung widerspricht der Unabhängigkeit der Faktoren und damit Guilfords S-I-Modell, das keinerlei Hinweise auf Zusammenhänge zwischen verschiedenen Operationsklassen enthält. Eine hierarchische Struktur wäre im übrigen auch für die Dimension der Denkprodukte zu erwarten; z.B. dürfte das Erkennen von Transformationen ohne die Fähigkeit, Relationen zu erkennen, nicht möglich sein. Das Modell

---

1) Dargestellt in Guilford and Christensen 1973

2) Ein solches Ergebnis wurde von McNemar bereits 1964 antizipiert, vgl. S. 879.

Guilfords trägt jedoch in seiner jetzigen Form der Wahrscheinlichkeit unterschiedlicher Beziehungen zwischen den 12o Fähigkeitsfaktoren keine Rechnung[1].

2.13 Kritische Einwände gegen die empirische Überprüfung der Theorie

Eine weitere Gruppe von kritischen Argumenten richtet sich gegen Guilfords Vorgehen bei der empirischen Überprüfung seines theoretischen Ansatzes.

Guilford (195o) beschreibt seine Untersuchungsmethode folgendermaßen: Er wählt einen Untersuchungsbereich aus, z.B. denjenigen der Gedächtnisfähigkeiten, und stellt Hypothesen auf über die Art der Faktoren in diesem Bereich. Die Hypothesenbildung schließt die Aufstellung von Alternativhypothesen über spezifische Merkmale der Faktoren ein und bildet damit die Grundlage für die Demonstration der vermuteten Faktoren mit Hilfe von bereits bekannten oder zu diesem Zweck entwickelten Testaufgaben. Die für eine Untersuchung zusammengestellte "Testbatterie" enthält immer einige "Referenztests", von denen bekannt ist, welche Faktoren sie messen. Auf diese Weise können eventuell Alternativhypothesen über den durch bestimmte Tests gemessenen Faktor ausgeschlossen werden. Die Testbatterie wird einer angemessen großen Stichprobe gegeben (nach Guilford 1964 scheint die Zahl der Probanden häufig etwas über 2oo zu liegen), die nach Geschlecht, Alter und Bildungsstand relativ homogen ist. Die Testergebnisse werden nach der Hauptkomponenten-Methode faktoranalysiert und - nach einer Methode von Cliff (1966) - orthogonal rotiert mit dem Ziel der Annäherung an eine "theoretical target matrix" (Guilford und Hoepfner 1969), die die jeweils zu untersuchenden hypothetischen Faktoren des S-I-Modells enthält. Kritische Einwände gegen Details dieses Vorgehens richten sich u.a. gegen die Homogenität der verwendeten Stichproben[2] und gegen die Subjektivität der Rotationsmethode, die darin besteht, daß die Hypothese über den durch einen Test erfaßten Faktor die endgültige Faktorenlösung mitbestimmt[3]. Eine Untersuchung von Horn (1967) mit 74 Zufallsvariablen (N = 3oo für jede der Variablen) zeigt, daß mit Hilfe einer solchen subjek-

---

1) Messick (1973) zeigt eine Möglichkeit auf , dem S-I-Modell durch Faktoren 2. und 3. Ordnung eine hierarchische Struktur zu verleihen.

2) Die Homogenität der Stichproben Guilfords wird von Anhängern eines g-Faktors dafür verantwortlich gemacht, daß in seinen Faktorenanalysen kein starker g-Faktor auftritt, s.z.B. Vernon 1969.

3) S.z.B. Guilford 1967b, S.199-2oo: "In the first rotational solution this test went with factor CMU, which did not seem reasonable..." "In a subsequent rotation, in which room was made for an additional factor and an axis was aimed at this test, it came out with a loading of +.55 on a factor that might be EMT."

tiven Rotationsmethode ("Procrustes Program" von Hurley und Cattell 1962, zitiert nach Horn 1967) interpretierbare Faktoren gefunden werden können; subjektive Rotationen im obigen Sinn führen also - vorsichtig ausgedrückt - nicht zwangsläufig zu einer Falsifizierung von unzutreffenden Untersuchungshypothesen. Einige Kritiker der Theorie Guilfords sind daher der Ansicht, daß die von Guilford bevorzugte faktorenanalytische Methode nicht geeignet ist, sein theoretisches Modell zu überprüfen[1].

Noch schwerwiegender als diese methodischen Einwände ist die Kritik an dem gesamten Forschungsansatz Guilfords, die von McNemar (1964) folgendermaßen zusammengefaßt wird: "The structure of intellect that requires 12o factors may very well lead ... us, to regard our fractionization and fragmentation of ability into more and more factors of less and less importance, as indicative of scatterbrainedness." (S. 872). Sowohl Eysenck (1967) als auch Vernon (1964) weisen darauf hin, daß es mit Hilfe der Faktorenanalyse und ganz spezifischer Tests möglich sei, Gruppenfaktoren in immer enger definierte Faktoren aufzusplittern und daß dieser "reductio ad absurdum of the whole approach"[2] durch die statistische Methode keine Grenze gesetzt werde. Die psychologische Bedeutung so spezieller Faktoren erscheint denn auch mehr als zweifelhaft; zur Charakterisierung einer bestimmten Fähigkeit wie z.B. DMT (divergente Produktion semantischer Transformationen) greift Guilford häufig auf eine kurze Beschreibung der Testaufgaben zurück, die zur Messung dieses Faktors entwickelt wurden[3].
Nach Wewetzer (1972) haben viele dieser Faktoren nur einen begrenzten und vor allem einen"labilen Status"; auf dem Kontinuum "spezifisch-generell" variieren sie erheblich (vgl. S. 34 - 35).

Von besonderem Interesse ist in diesem Zusammenhang eine Aufstellung der gesicherten Intelligenzfaktoren von Pawlik (1968, S. 334-362). Danach wurden von den etwa 1oo Faktoren, die Guilford und seine Mitarbeiter gefunden haben, nur 16 auch in Untersuchungen anderer Intelligenzforscher nachgewiesen.

---

1) Z.B. Carroll 1968, 1972; Horn and Knapp 1973; Wallach 197o

2) Eysenck 1967, S. 82

3) Z.B. Guilford 1967c, 197o[4], Guilford and Hoepfner 1966, Guilford and Hoepfner 1971

Die Interpretation seiner Faktoren als intellektuelle Fähigkeiten fordert ebenfalls Kritik heraus. Vernon (1969) weist darauf hin, daß Faktoren zunächst Klassifikationen von Tests sind, und wirft Guilford ebenso wie Thurstone vor, daß sie Faktoren geradezu als chemische Elemente des Verstandes betrachten, die in ihrer Gesamtheit alle wichtigen menschlichen Fähigkeiten ausmachen. In der Tat kann man sich bei der Lektüre Guilford'scher Publikationen oft des Eindrucks nicht erwehren, als wären seine 12o Faktoren für ihn nicht mehr theoretische Konstrukte, sondern hätten reale Existenz angenommen[1].

Diese Sichtweise ist umso erstaunlicher, als bisher kaum empirische Belege für die Vorhersage- oder wenigstens für die Übereinstimmungsgültigkeit der Tests vorliegen, da Guilfords nahezu ausschließlich verwendete Validierungsmethode in der Überprüfung der faktoriellen Validität besteht[2]. Er begründet dies damit, "daß es nur, nachdem wir die vielversprechenden Faktoren und die zugehörigen Meßmethoden bestimmt haben, gerechtfertigt ist, die Zeit schöpferischer Menschen mit Tests zu beanspruchen."[3] Wie die Entscheidung darüber, ob ein Faktor vielversprechend ist oder nicht, ohne die Verwendung von Außerkriterien getroffen werden kann, bleibt offen.

2.14 Kritische Einwände gegen Guilfords Aufgaben zur Erfassung kreativer Fähigkeiten

Die mangelnde äußere Validierung der Tests zur Erfassung kreativer Fähigkeiten wurde schon relativ früh kritisiert[4]. Trotzdem führten Guilford und seine Mitarbeiter bis in die siebziger Jahre eine ungeheure Anzahl von Untersuchungen durch, die alle das Ziel hatten, die 12o postulierten Faktoren seines S-I-Modells nachzuweisen. Da dies mit einer statistischen Methode geschah, die nicht zwangsläufig zu einer Falsifizierung unzutreffender Untersu-

---

1) S.z.B. Guilford 197o, S. 157: "First, there is not just one divergent-production ability or function; there are twenty-four of them, all more or less independent. This is not just theory; twenty-three of them have been demonstrated by factor analysis."

2) In dem Abschnitt "The validation of tests of creative talent" eines 1971 erschienen Artikels kann Guilford nur auf 3 Untersuchungen verweisen, in denen DP-Testergebnisse mit Außenkriterien korreliert wurden; Kreuzvalidierungen wurden nicht durchgeführt.

3) Vgl. Guilford 195o in Mühle und Schell (Hrsg.) 197o, S. 35

4) Z.B. von McNemar 1964, Vernon 1964, Yamamoto 1965a, aber auch von Wallach 197o

chungshypothesen führt (s. dazu S. 9/1o), ist die empirische Überprüfung der Theorie bestenfalls als unzureichend zu bezeichnen. Daß überdies die Validierung der dabei verwendeten Tests nur darin bestand, mit diesem angreifbaren statistischen Verfahren die faktorielle Reinheit der Testaufgaben zu überprüfen und nicht zusätzliche Untersuchungen zur äußeren Validität durchgeführt wurden, ist im Hinblick auf forschungsökonomische Gesichtspunkte unverständlich. Wenn nach 25-3o Jahren intensiver Forschungsarbeit alle oder fast alle postulierten Faktoren nachgewiesen sein werden und sich bei einer erst dann einsetzenden äußeren Validierung zeigen sollte, daß für Tests zur Messung so spezieller Fähigkeiten keine Außenkriterien gefunden werden können, weil alle realen kreativen Leistungen einen viel höheren Komplexitätsgrad aufweisen, daß also mit anderen Worten weniger spezifische Tests eine höhere Übereinstimmungs- und Vorhersagevalidität erzielen als die Guilford-Aufgaben, dann könnte Guilford der Vorwurf einer ungeheuren Verschwendung von Forschungskapazität gemacht werden.

Während diese Kritikpunkte und alle vorher dargestellten Überlegungen prinzipiell nicht nur die für kreatives Verhalten bedeutsamen Faktoren, sondern das gesamte theoretische Modell Guilfords betreffen, richtet sich ein letzter Einwand ausschließlich gegen Guilfords Kreativitätstests und die dahinterstehende Auffassung über den Einfluß motivationaler Faktoren auf die Testleistung.
Eine Reihe von Kreativitätsforschern ist der Ansicht, daß motivationale Variablen nicht nur beim Zustandekommen kreativer Leistungen eine wichtige Rolle spielen, sondern auch bei der Bearbeitung von Tests, die kognitive Voraussetzungen für kreatives Verhalten untersuchen. Dieser Auffassung wurde beispielsweise von Rookey[1)] durch Entwicklung eines Fragebogens zur Erfassung der "kreativen Tendenz", von Torrance durch die Konstruktion von Testaufgaben, die Neugier und Interesse der Kinder wecken, und von Wallach und Kogan durch eine entspannte, spielerische Testsituation ohne Zeitbegrenzung Rechnung getragen.
Guilford (1967c)ist zwar auch der Meinung, daß motivationale und Temperamentsvariablen "have some possibility of relevance as contributors to potential for creative production" (S. 282), aber er sieht offenbar keinen Anlaß, einen eventuellen Einfluß dieser Variablen auf die Testleistung zu berücksichtigen. Obwohl für die

1) Rookey 1971a, zitiert nach Rookey and Reardon 1973

meisten Aufgaben möglichst viele Lösungen verlangt werden und der Schwierigkeitsgrad z.T. recht hoch ist, wird nicht versucht, die Motivation der Probanden zu beeinflussen. Die Instruktionen sind knapp und sachlich und lassen keinerlei Versuch erkennen, diese Aufgaben von Intelligenztest-Items abzuheben. Noch erstaunlicher ist die z.T. überaus knappe Zeitbegrenzung (z.B. bei "Making Objects": aus vorgegebenen Figuren und Linien sind für 2 Teile mit je 9 Items 6 Minuten Testzeit vorgesehen, s. dazu Ulmann 1968, S. 81). Anläßlich einer kritischen Auseinandersetzung mit dem Kreativitätstest von Wallach und Kogan begründet Guilford (1971) sein von diesen beiden Autoren völlig abweichendes Vorgehen. Er ist der Ansicht, daß es für Kinder in fünften Klassen keinen Unterschied ausmacht, ob ein Test als Test oder als Spiel bezeichnet wird. Der durch die spielerische Testsituation eventuell erreichten Entspannung der Kinder sei durch die Einzeltestung und einen dauernd anwesenden und protokollierenden Testleiter entgegengearbeitet worden, so daß in Wirklichkeit die relative Anonymität der Gruppentestung von den Kindern eventuell sogar als angenehmer empfunden werde.
Während diese Argumentation einiges für sich hat, sind Guilfords Einwände gegen einen Wegfall der Zeitbegrenzung weniger überzeugend. Er hat in einer zusammen mit Christensen durchgeführten Untersuchung festgestellt, daß bei Tests zur Erfassung der Assoziationsflüssigkeit (also des Faktors "divergente Produktion semantischer Relationen") die während der ersten 2 Minuten gegebenen Antworten "bessere" Messungen dieses Faktors waren als die während der letzten 3 Minuten gegebenen. Außerdem seien Tests ohne Zeitbegrenzung unökonomisch, weil die Probanden dann sehr langsam arbeiten würden.

Während das letzte Argument nur unter der Voraussetzung einer unter beiden Bedingungen unveränderten Validität stichhaltig wäre und somit erst nach Vorliegen vergleichbarer Validitätsuntersuchungen für die Tests von Guilford und von Wallach und Kogan relevant werden dürfte, betrifft der Einwand, bei Wegfall der Zeitbegrenzung leide die faktorielle Validität von Testaufgaben, nur solche Verfahren, die die Erfassung reiner Faktoren anstreben. Wallach und Kogan wollten mit ihrem Kreativitätstest ein Verfahren schaffen, dessen Untertests untereinander hoch und mit Intelligenztestaufgaben nicht korrelieren. Daß ihnen das gelungen ist, zeigen nicht nur ihre

eigenen Untersuchungsergebnisse, sondern beispielsweise auch eine von Krause (1972) durchgeführte Faktorenanalyse, die einen klaren Kreativitätsfaktor erbrachte, auf dem die Untertests von Wallach und Kogan Ladungen zwischen .47 und .77 (Md=.54), die Untertests des Leistungsprüfsystems von Horn Ladungen von .06 bis .26 (Md=.13) hatten. Da Wallach und Kogan sich nie das Ziel gesetzt hatten, mit ihren 10 Testcores 10 Faktoren aus Guilfords S-I-Modell zu erfassen, dürfte sie der Vorwurf, keine reinen Faktoren sensu Guilford zu messen, nicht berühren. Immerhin besteht die Möglichkeit, daß sich durch den Wegfall der Zeitbegrenzung die äußere Validität des Tests von Wallach und Kogan erhöht; Guilford vermutet, daß Persönlichkeitsvariablen wie Ausdauer oder Höhe des Anspruchsniveaus im Verlauf der Testbearbeitung ins Spiel kommen. Da die Persistenz bei der Verfolgung selbstgesteckter Ziele eine wichtige Voraussetzung für reale kreative Leistungen ist, könnte gerade diese von Guilford kritisierte Vermischung von Fähigkeits- mit motivationalen Variablen zur Erhöhung der letztlich entscheidenden prädiktiven Validität des Tests von Wallach und Kogan beitragen.

Guilford zieht solche Gesichtspunkte überhaupt nicht in Erwägung. Er ist der Ansicht, daß "unless the test is logically or empirically shown to have a positive relation to one of the divergent-production or transformation ablilities of the SI model (including hypothesized auditory-figural and kinesthetic-figural DP abilities), its status as a special contributor to assessment of creative talent is to be questioned. The best kind of empirical check is to factor analyze the test, with appropriate controls."[1)]
Da Guilford auch den Remote Associates Test von Mednick allein aufgrund seiner Beziehungen zu Faktoren des S-I-Modells beurteilt[2)], entsteht der Eindruck, als sei er durch die ausschließliche Konzentration auf die kreativitätsrelevanten Faktoren seines theoretischen Modells und die alleinige Verwendung einer bestimmten Validierungsmethode blind geworden gegen andere Kreativitätstheorien und andere Möglichkeiten der Testentwicklung und -validierung. Seine Überzeugung, daß nur Untersuchungen, die sich an seinen Faktoren orientieren, die Kreativitätsforschung voranbringen können, läßt es wenig wahrscheinlich erscheinen, daß er berechtigte Einwände berücksichtigen und seinen Forschungsansatz revidieren oder zumindest durch Untersuchungen zur äußeren Validität seiner Tests ergänzen wird.

---

1) S. Guilford 1971, S. 83.
2) S. Guilford 1971, S. 83-84.

2.15 Schlußfolgerungen für die weitere Arbeit

Im Hinblick auf die Zielsetzung der vorliegenden Arbeit, ein für 9- bis 12jährige Schüler geeignetes Verfahren zur Erfassung kognitiver Voraussetzungen für kreatives Verhalten zu adaptieren oder zu entwickeln, erscheint es nicht angezeigt, von Guilfords theoretischem Bezugssystem auszugehen. Abgesehen von den besprochenen Schwächen und Nachteilen des S-I-Modells ist es Guilford und seinen Mitarbeitern bisher nicht gelungen, die Bedeutung der kreativtätsrelevanten Faktoren des Modells für diese Altersstufe nachzuweisen.[1] Außerdem würde ein Verfahren, das einen wesentlichen Teil dieser Faktoren erfaßt, für jüngere Schüler zu umfangreich.[2] Während es also weder möglich noch sinnvoll erscheint, alle 5 Produktklassen der divergenten Produktion und sämtliche Transformationsfähigkeiten zu messen, wird es weiterhin als wünschenswert betrachtet, ein Verfahren zu entwickeln, das Aufgaben aus einigen oder allen Inhaltsbereichen sensu Guilford umfaßt und dadurch mehrere unterschiedliche Begabungs- und/oder Interessenrichtungen berücksichtigt.

Vor allem jedoch scheint Guilfords Konzept des divergenten Denkens und seine Methode zur Messung der zugrundeliegenden Fähigkeiten eine brauchbare Grundlage für die Konstruktion des zu entwickelnden Tests zu bieten. Trotz der Schwierigkeiten, die bei der Auswertung und statistischen Verarbeitung der Testergebnisse auftreten, dürften Items mit freier Aufgabenbeantwortung die beste Methode zur Untersuchung kognitiver Voraussetzungen für kreatives Verhalten sein. Guilfords Methode der Testvalidierung wird aufgrund der dargestellten Kritikpunkte nur als eine von mehreren Validierungsmöglichkeiten betrachtet.

2.2 Die Torrance Tests of Creative Thinking (TTCT)

Ende der fünfziger Jahre begann Torrance, ursprünglich ein Anhänger des theoretischen Modells von Guilford[3], einige von Guilfords Tests für Kinder zu adaptieren. Zusätzlich entwickelte er - ausgehend von Veröffentlichungen über wissenschaftliche Entdeckungen und die Psychologie der Entdecker[4] - neue Tests. Nach vielfältigen

---

1) Nicht einmal alle nach der Theorie postulierten Faktoren konnten identifiziert werden, s. dazu Anmerkung 2, S. 5

2) Die in Anmerkung 3, Seite 3 erwähnte Untersuchung mit 4oo Schülern sechster Klassen dauerte z.B. 4 Stunden!

3) S. Goldman 1964, S. 4

4) Literaturangaben dazu s. Goldman 1964

Erprobungen und Verbesserungen legte Torrance 1966 eine Zusammenstellung dieser Aufgaben unter dem Namen "Torrance Tests of Creative Thinking" in 2 Parallelformen vor. Die Torrance-Tests sind nach seinen Angaben für Kinder und Jugendliche vom Kindergarten- bis zum College-Alter geeignet. Deswegen und wegen ihrer im Vergleich mit anderen Kreativitätstests relativ weit fortgeschrittenen Standardisierung dürften die TTCT inzwischen der am häufigsten verwendete Kreativitätstest geworden sein.
Die Testaufgaben selbst wurden so oft beschrieben[1], daß sie hier nicht noch einmal dargestellt werden sollen; dagegen soll im folgenden eine Zusammenfassung der theoretischen Fundierung, ein Überblick über die Ergebnisse von Reliabilitäts- und Validitätsuntersuchungen gegeben werden und es sollen die mit diesem Verfahren erzielten und für die vorliegende Arbeit relevanten Untersuchungsergebnisse referiert werden.

2.21 Theoretische Fundierung

Torrance hat nach seiner Auffassung Testaufgaben entwickelt, die "Modellfälle des kreativen Denkprozesses darstellen, wobei jedes Item verschiedene Arten des Denkens involviert und jedes etwas Einzigartiges zu den in der Entwicklung stehenden Aufgabenreihen beiträgt"[2]. Entsprechend dieser Absicht sind die Testitems nach Form und Inhalt sehr verschieden. Die seit 1966 im Handel erhältliche "Research Edition" gliedert sich in 2 unabhängig administrierbare Teile, die verbalen und die figuralen Tests. Erstere bestehen aus 2 rein verbalen Items ("Unusual Uses" und "Unusual Questions") und 3 verbal zu beantwortenden Items mit visuellen Aufgabenstimuli ("Ask-and-guess", "Product Improvement" und "Just Suppose"), letztere enthalten 3 zeichnerisch zu lösende Items ("Picture Construction", "Incomplete Figures" und "Repeated Figures").Die Aufgaben sind z.T. untergliedert; z.B. sollen bei "Ask-and-Guess" zuerst möglichst viele Fragen zu einem Bild gestellt werden, dann sollen möglichst viele Ursachen für die Begebenheit auf dem Bild gefunden und schließlich möglichst viele Konsequenzen der dargestellten Situation genannt werden.
Diese Aufgaben wurden nicht auf der Grundlage einer einheitlichen, umfassenden Kreativitätstheorie entwickelt. Torrance (1966) be-

1) Z.B. Goldman 1964, Krause 1972, Seiffge-Krenke 1974, Torrance 1968a und 1968b, Ulmann 1968

2) Vgl. Torrance 1968a, S. 334

schreibt das kreative Denken als "a process of becoming sensitive to problems, deficiencies, gaps in knowledge, missing elements, disharmonies, and so on; identifying the difficulty; searching for solutions, making guesses, or formulating hypotheses about the deficiencies; testing and retesting these hypotheses and possibly modifying and retesting them; and finally communicating the results." (S. 6). Diese Aufzählung der einzelnen Schritte eines Problemlösungsprozesses stellt bereits die gesamte theoretische Fundierung der Torrance Tests of Creative Thinking dar.
Torrances Vorgehen bei der Testkonstruktion war eklektisch; er schreibt dazu: "To insure content validity, a consistent and deliberate effort has been made to base the test stimuli, the test tasks, instructions, and scoring procedures on the best theory and research now available."[1] Angaben über die bei der Auswahl der Theorien und Forschungsergebnisse verwendeten Kriterien, die eventuell einige weitere Aufschlüsse über Torrances eigenes implizites Kreativitätskonzept geben könnten, werden nicht gemacht. Torrance behauptet lediglich, seine Tests stellten eine starke Abweichung von den faktorenanalytisch konstruierten Tests von Guilford und seinen Mitarbeitern dar. Er habe sich auch bemüht, in sein Verfahren andere Aufgaben als diejenigen von Wallach und Kogan einzubeziehen, nur die Forderung dieser beiden Autoren, "in den Testinstruktionen und in der Gestaltung etwas von dem Anreiz zur freien spielerischen Betätigung beizubehalten"[2], habe er zu verwirklichen versucht.
Im folgenden soll untersucht werden, ob und in welchem Ausmaß diese Zielsetzungen bei der Testkonstruktion verwirklicht wurden.
Zunächst muß festgestellt werden, daß die Instruktionen zu den einzelnen Aufgaben und die Testsituation vermutlich wenig Anreize zur "freien spielerischen Betätigung" enthalten. Die Aufgaben haben Zeitbegrenzung und die Kinder werden aufgefordert "Work as fast as you can without rushing".[3] Crockenberg (1972) weist in diesem Zusammenhang darauf hin, daß die intrinsische Motivation, die nach Ansicht vieler Kreativitätsforscher von besonderer Bedeutung für kreatives Verhalten ist, nicht nur durch den Zwang, eine nicht vollendete Aufgabe unerledigt zu lassen, beeinträchtigt wird, sondern auch dadurch, daß die Kinder sich mit einer Aufgabe, zu der

1) S. Torrance 1966, S. 24
2) Vgl. Torrance 1968a, S. 334
3) Torrance 1966, S. 5

ihnen nichts mehr einfällt, so lange beschäftigen müssen, bis die Testzeit abgelaufen ist. Auch die Aufforderung, an Lösungen zu denken, an die sonst niemand denkt, dürfte von den Kindern als Erschwernis der Aufgabe und nicht als Anreiz zu spielerischer Betätigung empfunden werden.
Die Absicht, andere Aufgaben als diejenigen von Wallach und Kogan in den Test aufzunehmen, wurde realisiert: der Torrance-Test enthält z.T. wesentlich komplexere Aufgaben und vor allem nicht nur verbal zu beantwortende Items, sondern auch solche, bei denen die Kinder malen müssen.
Die Behauptung, daß die TTCT eine starke Abweichung von den Guilford-Tests darstellen, trifft insofern zu, als Torrance seine Tests nicht mit Hilfe der Faktorenanalyse validierte. Abgesehen davon jedoch enthalten die TTCT Aufgaben von Guilford, die von Torrance nur für Kinder adaptiert wurden, und sie werden nach 4 Kriterien - Flüssigkeit, Flexibilität, Originalität und Elaboration[1] - ausgewertet, die in Guilfords Modell wichtige Faktoren der divergenten Produktion sind. Andere Faktoren, die von Guilford ebenfalls als Voraussetzungen für kreatives Verhalten betrachtet werden (vgl. S. 5), bleiben unberücksichtigt, eine Begründung für die Auswahl wird nicht gegeben. Die Tatsache, daß Torrance ganz verschiedene Aufgabentypen nach diesen 4 Kriterien auswertet, während Guilfords Faktoren unterschiedlich spezifisch im Hinblick auf die implizierten Produkt- und Inhaltsklassen sind[2], wird ebenso wenig besprochen oder gar begründet wie die Operationalisierung der Originalität, die bei Torrance ein Maß für die sta-

---

1) Flüssigkeit ist definiert durch die Anzahl der akzeptablen Lösungen, Flexibilität durch die Anzahl der verschiedenen Lösungskategorien, Originalität durch die statistische Seltenheit der Antworten und Elaboration durch die Anzahl ausschmückender Details.

2) Als Flüssigkeitsfaktoren werden die insgesamt 12 Faktoren der divergenten Produktion von Einheiten, Relationen und Systemen in allen 4 Inhaltsbereichen bezeichnet. Etwas weniger umfassend sind die Faktoren der spontanen Flexibilität und der Elaboration; beide Faktorengruppen wurden ebenfalls für alle 4 Inhaltsbereiche nachgewiesen, beziehen sich jedoch jeweils nur auf eine Produktkategorie, die spontane Flexibilität auf die Kategorie der Klassen, die Elaboration auf die Kategorie der Implikationen. Noch wesentlich spezifischer ist Guilfords Originalitätsfaktor, der nur eine Produktklasse - diejenige der Transformationen- und nur einen einzigen Inhaltsbereich - den semantischen - beinhaltet.

tistische Seltenheit der jeweiligen Antwort ist, während Guilford außerdem 2 weitere Originalitätsoperationalisierungen verwendet.[1)]

Torrance (1966) empfiehlt, die Flüssigkeits-, Flexibilitäts- und Originalitätsscores aller verbalen Items zu summieren (auf besonderen Wunsch könne auch ein verbaler Elaborationsscore ermittelt werden); bei den figuralen Tests werden Summenwerte für alle Kreativitätsmaße berechnet. In Einzelfällen kann auch aus diesen 3 bzw. 4 Werten ein verbaler und/oder figuraler Gesamtwert gebildet werden.
Eine Summation über verschiedene Aufgaben hinweg widerspricht völlig der erklärten Absicht des Testautors, daß jedes Item etwas Einzigartiges zu der Testbatterie beitragen solle (vgl. S. 16). Die Berechnung von verbalen und figuralen Gesamtwerten für Flüssigkeit, Flexibilität, Originalität und Elaboration wäre angemessen, wenn diese Maße untereinander unabhängig wären (keine oder geringe "Intratask-Konsistenz"), aber z.B. der Flüssigkeitsscore von Aufgabe 1 mit demselben Wert für die anderen Aufgaben hoch korrelieren würde (dasselbe gilt für die übrigen 3 Kreativitätsmaße), mit anderen Worten wenn die "Intertask-Konsistenz" hoch wäre.
Während eine hohe Intertask-Konsistenz nicht zu erwarten ist, wenn Torrance seine Absicht, heterogene Items zu konstruieren (s.o.), verwirklichen konnte, ist die für die Bildung der verschiedenen Summenwerte erforderliche niedrige Intratask-Konsistenz wegen der Abhängigkeit der 4 Kreativitätsmaße unwahrscheinlich. Wenn ein Kind innerhalb einer nicht zu knapp bemessenen Testzeit zu einem bestimmten Item nur Antworten innerhalb einer einzigen Antwortkategorie gibt, fällt nicht nur der Flexibilitätsscore schlecht aus, sondern mit hoher Wahrscheinlichkeit auch der Flüssigkeitsscore, weil es mehr Zeit beanspruchen dürfte, eine Antwortklasse auszuschöpfen als nach den jeweils geläufigsten Assoziationen die Klasse zu wechseln und auch da nur die Einfälle aufzuschreiben, die sich ohne langes Nachdenken einstellen, und sich dann einer weiteren Klasse zuzuwenden etc. Ebenso scheint die Häufigkeit statistisch seltener Einfälle mit der Gesamtzahl der Einfälle zu steigen.[2)]
Im Gegensatz dazu müßte zwischen Flüssigkeit und Elaboration eine gegenläufige Beziehung erwartet werden, da bei Aufgaben mit Zeitbegrenzung entweder eine geringe Zahl von Lösungen sehr detailliert

---

1) "Cleverness" als Güte der Antwort (nur durch Expertenrating zu bewerten) und "remoteness" im Sinn von Entlegenheit der Assoziationen.

2) Dieser Zusammenhang kann aus der assoziationstheoretischen Kreativitätstheorie von Mednick abgeleitet werden (s. dazu Wallach and Kogan 1965, S. 13-2o).

ausgearbeitet und ausgeschmückt oder eine größere Zahl von Lösungen relativ wenig elaboriert werden.[1)]
Ob Torrance solche oder ähnliche Überlegungen angestellt hat, ist seinen Veröffentlichungen zur Testkonstruktion nicht zu entnehmen; er berichtet lediglich von Versuchen, außer den 4 genannten Auswertungskriterien noch eine ganze Reihe weiterer zu erproben ("communication, inventivlevel, constructiveness, penetration, resistance to closure and the like").[2)]
Bevor jedoch Schlußfolgerungen daraus gezogen werden konnten, mußten diese Versuche wegen des damit verbundenen ungeheuren Zeitaufwandes aufgegeben werden.
Die relativ ausführliche Darstellung der Auswertungsprozedur der TTCT sollte zeigen, daß Torrances eklektizistisches Vorgehen zu einer Reihe von nicht begründeten und willkürlich erscheinenden Entscheidungen führte. Als noch schwerwiegender wird der offensichtliche Widerspruch zwischen dem Prinzip der Testkonstruktion, Items zusammenzustellen, von denen jedes etwas Einzigartiges mißt, und demjenigen der Testauswertung, Summenwerte über die einzelnen Aufgaben hinweg zu bilden, betrachtet. Auch ohne eine zugrundeliegende einheitliche Kreativitätstheorie und bei einem eklektizistischen Vorgehen dürften solche Inkonsistenzen nicht unvermeidbar sein.

Im folgenden sollen Torrances theoretische Annahmen über die Beziehung zwischen Kreativität und Intelligenz erörtert werden. Diese Annahmen werden allerdings nirgends detailliert dargelegt, sondern müssen aus einer ganzen Reihe verschiedener Quellen erschlossen werden.
Torrance (1962a)ist der Ansicht, daß es sich bei den durch traditionelle Intelligenztests gemessenen Fähigkeiten und denjenigen, die die Voraussetzung für kreatives Verhalten sind, um zwei eindeutig unterscheidbare Dimensionen handelt. Diese Überzeugung kommt in seinen langjährigen intensiven Bemühungen um die Konstruktion eines Kreativitätstests zum Ausdruck, sie spiegelt sich in Äußerungen wie z.B. derjenigen, daß bei alleiniger Verwendung eines Intelligenztests 7o Prozent der hochkreativen Schüler übersehen werden[3)] und

---

1) S. Krause 1972, S. 8o
2) S. Torrance 1964a, S. 129
3) Torrance 1959, zitiert nach Ulmann 1968, S. 111

sie zeigt sich indirekt in zahlreichen Veröffentlichungen mit pädagogischer Zielsetzung, in denen Torrance Lehrer und Eltern über kreativitätsfördernde Erziehungsmaßnahmen aufzuklären versucht, wobei er immer wieder auch darauf hinweist, wie leicht kreative Schüler - im Gegensatz zu intelligenten - unentdeckt bleiben und wie leicht ihre schöpferischen Fähigkeiten infolge mangelnder Förderung verkümmern können.[1] Schließlich äußert sie sich auch in einer etwas einseitigen Berichterstattung über die Höhe der Korrelationen, die für den Zusammenhang zwischen IQ und den Ergebnissen der TTCT gefunden wurden. Torrance (1967a) gibt an, daß er selbst immer niedrige positive und manchmal statistisch signifikante Korrelationen gefunden habe. Eine Zusammenstellung von 178 Korrelationskoeffizienten für den Zusammenhang zwischen Intelligenztestergebnissen und einem Gesamtwert der TTCT aus verschiedenen Untersuchungen mit einem Median-Koeffizienten von .2o scheint seine Resultate zu bestätigen. Nicht dargestellt werden die schwerwiegenden methodischen Einwände gegen die meisten früheren Korrelationsuntersuchungen, die häufig mit Stichproben mit eingeschränkter Intelligenzstreuung durchgeführt wurden (überwiegend handelte es sich um Gruppen von hochintelligenten Schülern), nach Inhalt und Qualität völlig unterschiedliche Intelligenztests verwendeten und die relativ niedrige Reliabilität der Kreativitätstests nicht berücksichtigten.[2] 1965 führte Yamamoto eine Untersuchung mit den TTCT durch, die diese methodischen Einwände berücksichtigte, indem eine Korrektur wegen Streuungseinschränkung und eine weitere wegen Unreliabilität der Kreativitätstestergebnisse durchgeführt wurde.[3] Für 2 relativ verschiedene Stichproben von Fünftkläßlern(N_1= 461, N_2= 827) erhielt Yamamoto folgende Korrelationen : r_1=.33, r_2=.39 vor der Korrektur; r_1=.51, r_2=.54 nach beiden Korrekturen. Für zwei Unterstichproben aus den Schülern mit Intelligenzquotienten zwischen 11o und 13o sind dagegen die unkorrigierten Koeffizienten .o8 bzw. -.o9.

Torrances Zusammenstellung von Korrelationskoeffizienten enthält zwar einige Koeffizienten über .5o, aber der Aritkel enthält keinerlei Hinweise darauf, daß diese hohen Werte vermutlich zuverlässiger sind; der Autor behauptet im Gegenteil, die zentralen Tendenzen

---

1) Z.B. Torrance 1961, 1962a, 1963a, 1964b, 1965a, 1965b, 1967a, 1967b, 197o, 1971a, 1972a, 1972b, 1974

2) Z.B. Cureton 1964, zitiert nach Yamamoto 1965b, Marsh 1964, Ripple and May 1962

3) Als Reliabilitätsschätzung diente ein für 7o Schüler aus einer der beiden Stichproben ermittelter Retest-Reliabilitätskoeffizient von .79 (Zeitintervall: 8 Wochen). Da die TTCT in Parallelformen vorliegen, dürfte dies ein angemessenes Vorgehen sein. (Yamamoto 1965 b)

dieser Untersuchungen stimmten mit denjenigen seiner eigenen überein.

Eine Annahme von Torrance (1962a) über den unterschiedlichen Zusammenhang zwischen Kreativität und Schulleistung auf verschiedenen Intelligenzstufen wurde unter dem Namen "Schwellenhypothese der Intelligenz" bekannt. Torrance nahm an, daß ab einer bestimmten Intelligenzhöhe die Korrelation zwischen IQ und Kreativitätstestergebnissen völlig verschwindet. Diesen Schwellenwert vermutet er bei etwa 12o IQ-Punkten. Ab diesem Wert trägt die Intelligenz nichts mehr zu einer Steigerung der Schulleistungen bei, wohl aber die Kreativität. Damit wird das Phänomen der Overachiever, also der Schüler mit in Relation zu ihrer Intelligenz erwartungswidrig hohen Schulleistungen, durch die hohe Kreativität dieser Schüler erklärt.

Offen bleibt nun noch die Frage, welche Rolle die Intelligenz nach Torrance beim Zustandekommen überragender kreativer Leistungen spielt. Es ist anzunehmen, daß Torrance auch ein Mindestmaß an Intelligenz als Voraussetzung für reale kreative Leistungen betrachtet, diese aber nicht so hoch sein muß, daß zukünftige Erfinder, Entdecker, Künstler o.ä. bereits in der Schule aufgrund ihrer überragenden Intelligenz identifiziert werden könnten. Die von ihm postulierte Intelligenzschwelle bei einem IQ von etwa 12o[1] dürfte für Torrance auch bei der Vorhersage von kreativen Leistungen Erwachsener eine Rolle spielen. Das Problem der Vorhersage sozial anerkannter kreativer Leistungen ist für Torrance allerdings nicht so zentral wie beispielsweise für Guilford, weil Torrance kreativem Verhalten auch andere Funktionen als die Hervorbringung gesellschaftlich relevanter schöpferischer Leistungen zuerkennt. In seinen pädagogisch motivierten Veröffentlichungen klingt immer wieder die Überzeugung durch, daß eine Unterdrückung kreativer Fähigkeiten zu Störungen in der Persönlichkeitsentwicklung führen könne, was impliziert, daß kreatives Verhalten nach seiner Meinung eine psychohygienische Funktion erfüllen kann.[2]

Als letzter Punkt soll Torrances Ansicht über die Beeinflußbarkeit kreativer Fähigkeiten durch Umwelteinflüsse besprochen werden. Ebenso wie seine Annahmen über die Beziehung zwischen Kreativität und Intelligenz hat sie die Testkonstruktion selbst vermutlich nicht beeinflußt, aber sie wirkte sich bei einer ganzen Reihe von Validierungsversuchen aus.

Aus den vorangegangenen Ausführungen über die von Torrance vermuteten Zusammenhänge zwischen Kreativität und Intelligenz dürfte schon

---

1) S. Torrance 1962a, S. 63

2) S. z.B. Torrance 1962a, Kapitel 7

deutlich geworden sein, daß er kreatives Verhalten für stark umweltabhängig hält. Zahlreiche Veröffentlichungen mit Anregungen und Vorschlägen für kreativitätsförderndes Lehrerverhalten und kreative, stimulierende Unterrichtsgestaltung belegen dies (vgl. Anmerkung 1, S. 21) und seine interkulturellen Untersuchungen über Diskontinuitäten in der Entwicklung der Originalität bestätigen seine Ansichten (vgl. Torrance 1963b).In einer Erwiderung auf eine Kritik an seiner Verwendung des Begriffs "kreativ" kommt seine Überzeugung der Beeinflußbarkeit kreativen Verhaltens durch äußere Einflüsse ebenfalls zum Ausdruck: Torrance schreibt, daß ein hoher Punktwert in seinem Test die Wahrscheinlichkeit erhöhe, daß das betreffende Individuum sich kreativ verhält; der Besitz dieser Fähigkeiten garantiere natürlich kreative Leistungen ebensowenig wie ein hohes Intelligenztestergebnis intelligentes Verhalten garantiere (dieser Vergleich täuscht allerdings darüber hinweg, daß Torrance im Hinblick auf die Entfaltung des kreativen Potentials viel mehr umweltabhängige, vor allem erziehungsbedingte Hindernisse sieht als für die Intelligenzentwicklung).

Ob es Torrance gelungen ist, auf der Grundlage einiger Theorieteile, die er ohne nähere Begründung und ohne Bemühungen zur Integration aus der vorliegenden Kreativitätsliteratur auswählte, einen reliablen und vor allem validen Kreativitätstest zu entwickeln, soll im folgenden untersucht werden. Ein eklektizistisches und sogar ein atheoretisches Vorgehen bei der Testkonstruktion erlauben nicht zwingend den Schluß, daß der Test nicht valide sein kann. Schließlich kann die Übereinstimmungs- oder sogar die Vorhersagevalidität durch Korrelation der Testergebnisse mit bestimmten Verhaltensweisen außerhalb der Testsituation nachgewiesen werden, wobei nach Michel (1964, S. 49) "dieser Zusammenhang nicht psychologisch evident oder aufgeklärt zu sein braucht."[1)]
Auf dem Gebiet der Kreativitätsforschung begegnet diese Validierung durch Korrelationsschluß allerdings erheblichen Schwierigkeiten; stellt die Auswahl und quantifizierende Aufbereitung eines Kriteriumsverhaltens schon bei Erwachsenen ein Problem dar, so potenziert sich dieses bei Untersuchungen mit Schulkindern zu einer fast unlösbaren Aufgabe.

---

1) Zur Kritik an einem solchen "korrelationalen Validitätswissen" s. Kaminski 197o, vor allem S. 372-383

2.22 Testgütekriterien

Die folgende Darstellung beschränkt sich auf die 3 "Hauptgütekriterien" nach Lienert (1967[2]), also auf Objektivität, Reliabilität und Validität.

2.221 Objektivität

Da Torrance (1966) eindeutige und detaillierte Anweisungen für die Testdurchführung gibt, dürfte die Durchführungsobjektivität gewährleistet sein.
Seine Hinweise für die Interpretation der 4 verbalen und 4 figuralen Gesamtscores sind zwar ebenfalls relativ ausführlich, aber weniger eindeutig, was z.B. aus der häufigen Verwendung von Formulierungen wie "im allgemeinen", "in manchen Fällen" u.ä. geschlossen wird. Außerdem gibt Torrance zu, daß für die Interpretation einzelner Gesamtscores noch nicht genügend Erfahrungen vorliegen.
Die Auswertungsobjektivität stellt bei Kreativitätstests wie bei allen Verfahren mit freier Aufgabenbeantwortung ein besonderes Problem dar, das dadurch noch vergrößert wird, daß die Richtig-Falsch-Dichotomie kein geeignetes Bewertungskriterium für Antworten ist, deren Kreativität beurteilt werden soll.
Torrance hat allerdings sehr detaillierte Auswertungsrichtlinien für jede seiner Aufgaben und für alle 4 Auswertungskriterien zusammengestellt, so daß die Auswertungsobjektivität seines Verfahrens im Vergleich zu anderen Kreativitätstests relativ hoch sein dürfte. Außerdem führt er ein Auswerter-Training durch, das erst als abgeschlossen betrachtet wird, wenn die Korrelationen zwischen den Auswertungsergebnissen des unerfahrenen und denjenigen eines erfahrenen Auswerters mindestens .90 betragen.
Torrance (1966) gibt in einer Tabelle die Ergebnisse einer Untersuchung zur Auswertungsobjektivität der TTCT wieder, in der untrainierte Lehrer jeweils 25-35 Testhefte einer bestimmten Klassenstufe auswerteten. Die 30 Korrelationskoeffizienten für die 4 Gesamtwerte der figuralen und die 3 Gesamtwerte (ohne Elaboration) der verbalen Form verschiedener Klassenstufen (überwiegend Elementarschüler) liegen zwischen .66 und .99, der Median ist .955. Am objektivsten wird nach Flüssigkeit ausgewertet, es folgen Flexibilität und Elaboration; für die Originalitätsauswertung ergeben sich die niedrigsten Koeffizienten.

Die Auswertungsobjektivität ist für die verbalen Maße etwas höher als für die figuralen. Andere Untersucher fanden vergleichbare Korrelationen.[1]).

Die Auswertungsobjektivitäts-Koeffizienten sind für einen Test mit freier Aufgabenbeantwortung sehr hoch und dürften von allen anderen Kreativitätstests - abgesehen von Verfahren, die eine richtige Antwort verlangen wie z.B. der Remote Associates Test von Mednick - kaum erreicht werden.

2.222 Reliabilität

In diesem Abschnitt werden die Retest-Reliabilitäten der TTCT, ihre interne Konsistenz und die mit letzterer verbundene Frage der Inter-versus Intratask-Konsistenz (vgl. S. 19) besprochen.

Retest-Reliabilität

Im Hinblick auf die Retest-Reliabilität ist Torrance (1966, 1967a) der Ansicht, daß motivationale Faktoren die Stabilität von Kreativitätstestergebnissen stärker beeinflussen als die Ergebnisse von Persönlichkeits-, Intelligenz- und Leistungstests, da nur bei ersteren von den Probanden verlangt würde, die Antworten selbst zu produzieren. Da bei Erwachsenen eher mit einer konsistenten Motivation gerechnet werden könne als bei Kindern, seien die Retest-Reliabilitäten bei ersteren höher als bei letzteren.

Da motivationale Faktoren u.a. durch die Art der Testsituation beeinflußt werden, kann die Retest-Reliabilität bereits durch kleinere Änderungen der Testbedingungen herabgesetzt werden; Crockenberg (1972) äußert die Vermutung, daß die Retest-Reliabilität eine Funktion der Ähnlichkeit der beiden Testsituationen sei, eine Ansicht, die in dieser Ausschließlichkeit nicht haltbar sein dürfte.

---

1) Z. B. Barabasz (1969):.91-.99 (N=41, vermutlich Zweitkläßler).
Ingersoll and Snowman (1974): um .92 (bei Fünft- und Sechstkläßlern, keine weiteren Angaben).
Madaus (1967): .87-.99 für die 4 Kreativitätsmaße bei einzelnen verbalen und figuralen Tests, der Median für Flüssigkeit ist .985, für Flexibilität .9o, für Originalität .95, für Elaboration .94 (N=5o, 9.und 1o. Klasse high school; bei den Koeffizienten handelt es sich um Mittelwerte der Auswertungsobjektivitätskoeffizienten für 4 unabhängige Auswerter).
Wodtke (1964): .91-.99, Md.= .98 für den Gesamtwert aus allen verbalen und figuralen Tests; .92-1.oo, Md.= .95 für Subtest-Gesamtscores (N=65, 2.-5. Klassen).
Yamamoto (1965b):.99 für den Gesamtwert aus allen verbalen und figuralen Tests (N=64, Fünftkläßler, 2 Auswerter); .97-.99 für den gleichen Gesamtwert (N=76, Fünftkläßler, 4 Auswerter).

Torrance (1966) referiert die meisten der zu diesem Zeitpunkt vorliegenden Retest-Reliabilitäten für die 4 figuralen und 3 verbalen (ohne Elaboration) Gesamtscores; bei den Stichproben handelt es sich um Elementarschüler 4., 5. und 6. Klassen und College-Studenten; das Zeitintervall zwischen den beiden Testungen betrug zwischen 1-2 Wochen und 3 Jahren.
Die Reliabilitätskoeffizienten liegen zwischen .35 und .93, der Median ist .73.
Die Retest-Reliabilitäten für Gesamtwerte aus allen verbalen und figuralen Maßen,die nur für College-Studenten vorliegen (Zeitintervalle zwischen 1o Wochen und 3 Jahren) liegen zwischen .61 und .97 mit einem Median von .815.
Entsprechend niedriger liegen die von Torrance berichteten Werte für einzelne Aufgaben, sie betragen .47-.89 mit einem Median von .69; die Koeffizienten für die verbalen Items sind wie bei der Auswertungsobjektivität höher als diejenigen für die figuralen. Beim Vergleich der 4 Auswertungskriterien schneiden nach Torrance (1966) Flüssigkeits- und Flexibilitätsscores am besten ab, was aufgrund seiner Tabelle (vgl. S. 21) nur für die Flüssigkeitswerte bestätigt werden kann.
Nach 1966 wurden mehrere weitere Untersuchungen zur Retest-Reliabilität der TTCT durchgeführt[1)], deren Ergebnisse folgendermaßen

1) Autor(en)	Zeitintervall	Stichprobe	Reliabilitätskoeffizienten
Torrance and Aliotti 1969*)	1 Woche	118 5.Kläßler	.71-.93 (ohne nähere Angaben*)
Mackler and Shontz 1967	2 und 4 Wochen	4o Studentinnen	.47-.89, Md =.7o (ohne nähere Angaben)
Dewing 197oa	6 Wochen	394 7.Kläßler	.545-.864,Md=.675 für figurale Flüssigkeit; .297-.734,Md=.56o für figurale Originalität; .236-.728,Md=.476 für verbale Flüssigkeit; .o57-.665,Md=.38o für verbale Originalität.
Yamamoto and Chimbidis 1966	8 Wochen	7o 4.-6.Kläßler	.79 (Gesamtwert für alle verbalen und figuralen Maße)
Barabasz 1969	6 Monate	41 2.-und 3. Kläßler	.o47-.77o für die 4 figuralen Kreativitätsmaße, .53o für den figuralen Gesamtwert

Fortsetzung der Fußnote von Seite 26

Autor(en)	Zeitintervall	Stichprobe	Reliabilitätskoeffizienten
Ingersoll and Snowman 1974	6 Monate	5.-und 6. Kläßler (keine weiteren Angaben)	.3o-.67,Md=.5o
	12 Monate		.23-.6o,Md=.34
	18 Monate		.o8-.42,Md=.22
Haddon and Lytton 1971	4 Jahre	148 15jährige Sekundarschüler	.331-.621,Md=.495 für die verbalen und figuralen Gesamtwerte und die aus beiden berechneten Gesamt-Gesamtwerte
Cropley and Clapson 1971*)	5 Jahre	11o 12. Kläßler	.33-.58 (ohne nähere Angaben*)

*) zitiert nach Ingersoll and Snowman 1974

zusammengefaßt werden können: Die berichteten Koeffizienten streuen von .o47 bis .93; höhere Koeffizienten ergeben sich für kürzere Zwischenräume zwischen erster und 2. Testung, für Gesamtwerte aus allen verbalen und figuralen Maßen (im Vergleich mit den 4 figuralen und 3 verbalen Kreativitätsmaßen) und für ältere Probandengruppen. Die Untersuchung von Wodtke (1964), der über 5oo[1)] Schüler 2., 3., 4. und 5. Klassen nach 6 Monaten mit der Parallelform der TTCT erneut testete und die Ergebnisse sowohl für die einzelnen Kreativitätsmaße und den Gesamtwert jeder Aufgabe als auch für die 4 Gesamtwerte der figuralen und verbalen Tests sowie für einen figuralen und einen verbalen Gesamtwert aus den 4 Maßen darstellt, erlaubt einige zusätzliche Angaben: Die Retest-Reliabilitäten der einzelnen Items sind z.T. sehr unterschiedlich (z.B. Gesamtwert für "Picture Construction" r_{tt}=.o5, für "Circles" r_{tt}=.59, für "Unusual Uses" r_{tt}= .43, für "Ask-and-Guess" r_{tt}=.66, alle Koeffizienten für Schüler fünfter Klassen).
Die geringere Reliabilität der figuralen Tests geht zu Lasten der sehr unreliablen Aufgaben "Picture Construction" und "Figure Completion", für die nur wenige Koeffizienten über .3o (höchster Wert: r_{tt}=.36!) gefunden wurden.

1) Schwankendes N auf allen Klassenstufen wegen fehlender Daten für einzelne Variablen

Während die Ergebnisse von der 2. zur 3. Klasse keine Reliabilitätsverbesserung zeigen, werden die Koeffizienten von der 3. zur 4. und von der 4. zur 5. Klasse deutlich höher, erreichen jedoch nur für den Gesamtwert aus allen verbalen und figuralen Maßen und für den Gesamtwert aller verbalen Aufgaben Werte über .7o und dies nur in den 5. Klassen. Wodtke schließt aus seinen Ergebnissen, daß wegen der niedrigen Retest-Reliabilitäten, die nicht auf eine geringe Auswertungsobjektivität zurückgeführt werden können (vgl. Anmerkung 1, S.25)eine Verwendung der TTCT bei Elementarschülern nur für Forschungszwecke vertretbar sei.
Die unbefriedigenden Ergebnisse der Untersuchungen zur Retest-Reliabilität der TTCT werfen schwerwiegende Probleme hinsichtlich der Validität dieses Verfahrens auf. Wenn bereits geringfügige Änderungen der Testsituation oder auf andere Weise verursachte Motivationsänderungen zu so stark abweichenden Ergebnissen bei einer Testwiederholung führen und die Erklärung dafür nicht in Langeweile und Übersättigung gesucht werden kann, da mit einer Parallelform aus inhaltlich abweichenden Aufgaben getestet wird, dann sind die TTCT nicht geeignet, die zeitlich überdauernde Persönlichkeitsdimension, die eine wichtige Voraussetzung für kreatives Verhalten darstellt, zu messen. Ob die Retest-Reliabilität durch eine Änderung der Testsituation nach den Vorschlägen von Wallach und Kogan (1965), also durch Individualtestung in entspannter, spielähnlicher Situation ohne Zeitbegrenzung, erhöht werden könnte, kann nicht entschieden werden, da für das Verfahren von Wallach und Kogan, das keine Parallelform besitzt, keine vergleichbaren Ergebnisse vorliegen.[1] Einen Hinweis in dieser Richtung liefert eventuell eine Arbeit von Mondfrans, Feldhusen, Treffinger und Ferris (1971), in der Auswirkungen verschiedener Testbedingungen auf die Interkorrelationen der 3 verbalen und 4 figuralen Kreativitätsmaße der TTCT untersucht wurden; am höchsten interkorrelierten die 7 Maße in der Testsituation, die den Schülern erlaubte, die Testhefte für 4 Tage mit nach Hause zu nehmen und dort - ohne fremde Hilfe - daran zu arbeiten. Eine spielähnliche Situation, in der die Schüler die Zeitbegrenzung angeblich

---

1) Angaben zur Äquivalenz der Formen A und B konnten weder in dem sonst recht ausführlichen Manual von Torrance (1966) noch in einigen anderen Veröffentlichungen über die TTCT (Goldman 1964, Torrance 1964a,1967a,1968a)gefunden werden.

nicht merkten, führte dagegen zu niedrigeren Interkorrelationen der 7 Kreativitätsmaße als die von Torrance (1966) vorgeschriebene Standardprozedur. Daß der Wegfall der Zeitbegrenzung ein wichtiger Faktor ist, legt auch eine Untersuchung von Torrance (1969) nahe, nach der die Bearbeitung einer Testaufgabe während eines ganzen Tages zuhause zu valideren Ergebnissen führte als eine Gruppentestung mit Zeitbegrenzung. Torrance selbst fordert aufgrund dieser Ergebnisse mehr Untersuchungen mit Kreativitätstests ohne Zeitbegrenzung.

Interne Konsistenz

Die Frage der internen Konsistenz ist für die TTCT weniger kritisch als diejenige nach der Retest-Reliabilität, weil Torrance für sein Verfahren Aufgaben zusammenstellte, die "call into play different parts of a universe of abilities that may legitimately be conceptualized as "creative thinking abilities"."[1] Aus diesem Grund kann er die Kritik anderer Autoren[2] wegen der relativ niedrigen Interkorrelationen seiner Aufgaben als irrelevant zurückweisen. Nicht verteidigen kann er dann allerdings die von ihm empfohlene Summierung der Flüssigkeits-, Flexibilitätsscores etc. über verschiedene Aufgaben hinweg (vgl. S. 19). Zu den Interkorrelationen der 4 Kreativitätsmaße über verschiedene Aufgaben hinweg scheinen von Torrance selbst keine Angaben vorzuliegen, es gibt jedoch mindestens 2 Untersuchungen, denen einige Angaben zur internen Konsistenz zu entnehmen sind. Harvey, Hoffmeister, Coates und White (197o) berechneten für 64 Lehrer die Korrelationen der Flüssigkeits-, Flexibilitäts- und Originalitätsscores über alle verbalen Aufgaben hinweg; die Flüssigkeitsscores korrelieren zwischen .22 bis .73 (Md=.52), die Flexibilitätswerte erbringen Koeffizienten zwischen .18 und .59 (Md=.28) und für die Originalitätsscores ergaben sich Korrelationen zwischen .o2 und .55 (Md=.26).

Mackler und Shontz (1967) testeten 4o Studentinnen mit 4 Kreativitätsaufgaben, darunter 3 von Torrance (Ask-and-Guess, Unusual Uses und Repeated Figures); die Interkorrelationen der Flüssigkeitsmaße lagen zwischen .31 und.55, diejenigen der Flexibilitätsscores zwischen -.o1 und .26 und der als einziger berechnete Koeffizient für

---

1) Torrance 1966, S. 24

2) Thorndike 1962, Vernon 1964, Wallach and Kogan 1965

die Korrelation der Originalitätsscores von Unusual Uses und Repeated Figures war .33. O'Bryan und Mac Arthur (1969), die eine Untersuchung mit 85 neunjährigen Elementarschülern durchführten, teilen zwar nicht die Korrelationen zwischen den Kreativitätsmaßen verschiedener Aufgaben mit, aber sie stellen die Ergebnisse einer Faktorenanalyse dar, aus denen sie folgenden Schluß ziehen: "The factor analysis of the Torrance Tests did not demonstrate a relation between subdivisions of a given test and similarly named subdivisions of another test. Rather, the parts of each test appear to be strongly related. It was clearly not advisable to recombine the Torrance variables under such headings as flexibility, fluency, etc." (S. 39)

Die Bildung eines verbalen und eines figuralen Gesamtscores setzt auch hohe Korrelationen zwischen verschiedenen Kreativitätsmaßen verschiedener Aufgaben voraus. Diese Koeffizienten, die nur von Mackler und Shontz angegeben werden, liegen zwischen .11 und .47 mit einem Median von .28.

Diese Ergebnisse belegen trotz der geringen Versuchspersonenzahlen ausreichend die relativ niedrige interne Konsistenz der Aufgaben der TTCT.[1] Als letztes wird die interne Konsistenz - z.T. auch als Intertask-Konsistenz bezeichnet - der Interkorrelation der 3 bzw. 4 Kreativitätsmaße innerhalb jeder einzelnen Aufgabe - also der Intratask-Konsistenz - gegenübergestellt. Letztere müßte sehr gering und auf jeden Fall deutlich niedriger sein als erstere, wenn jede Aufgabe nach diesen 3 bzw. 4 Kriterien ausgewertet und die Scores über die verbalen bzw. figuralen Aufgaben aufsummiert werden sollen (vgl. S. 19).

---

1) Obwohl von Torrance nicht empfohlen, bilden manche Autoren einen Gesamtwert aus den Gesamtscores der verbalen und figuralen Tests. Dies ist ein völlig unangemessenes Vorgehen, da die 4 figuralen Maße mit den 3 verbalen noch wesentlich niedriger korrelieren (zwischen .23 und .52, Md=.395; die Daten wurden an 608 Sechstkläßlern von Cicirelli 1964 - zitiert nach Torrance 1966 - erhoben) als die 3 verbalen Maße untereinander (.79, .80, .74) und als die 4 figuralen Maße untereinander (.18 bis .77, Md=.50). Madaus (1967) untersuchte an 609 High School Sophomores ebenfalls die Korrelationen der verbalen und figuralen Kreativitätsmaße und fand in etwa die gleichen Ergebnisse: die verbalen Maße korrelierten zwischen .73 und .94 (mit einem Median von .85) untereinander, mit den figuralen Maßen jedoch nur zwischen .24 und .41 (Md=.33), während die figuralen Maße zwischen .46 und .87 (Md=.68) interkorrelierten.

Das Gegenteil ist der Fall: Die Intratask-Konsistenzen sind durchweg höher als die Intertask-Konsistenzen.[1] Insbesondere korrelieren die Flüssigkeits-, Flexibilitäts- und Originalitätsscores jeder einzelnen Aufgabe recht hoch untereinander[2]; Butcher (1973) folgert aus britischen Untersuchungen, die u.a. den Zusammenhang zwischen diesen 3 Maßen überprüfen, daß es wegen der hohen Interkorrelationen für die meisten Zwecke unerheblich sei, welches Maß verwendet wird. Hargreaves und Bolton (1972), die Aufgaben von Torrance, Guilford und Wallach und Kogan nach Flüssigkeit, Flexibilität, Originalität und z.T. nach Elaboration auswerteten und faktoranalysierten, schließen aus ihren Ergebnissen: "... the time and effort required to calculate anything more than Fluency scores on divergent tests do not justify the small amount of extra information gained." (S. 46o) Die erwartete negative Beziehung zwischen Elaboration und Flüssigkeit bei Aufgaben mit Zeitbegrenzung (vgl. S. 19) wurde nicht gefunden, wahrscheinlich ist die Testzeit für die meisten Kinder tatsächlich ausreichend bemessen[3]. Crockenberg (1972) folgert allerdings aus den Ergebnissen zweier größerer Untersuchungen mit Schulkindern, daß durch die Elaborationsscores etwas anderes gemessen werde als durch die anderen 3 Kreativitätsmaße, da sie die niedrigsten Korrelationen mit den anderen Maßen aufweisen. Diese Folgerung wird durch die Untersuchungsergebnisse von Seiffge-Krenke (1974) auch für bundesdeutsche Schüler bestätigt.

Es ist erstaunlich, daß die dargestellten Zusammenhänge zwischen den 4 Kreativitätsmaßen, die z.T. a priori erwartet werden konnten und die 1964 bereits durch 2 Untersuchungen bestätigt wurden[4], von Torrance in seinen Auswertungsrichtlinien nicht berücksichtigt werden. Drei der 4 Kreativitätsmaße besitzen eine hohe Intratask-Konsistenz, die Intertask-Konsistenz ist erheblich niedriger, aber dieser Umstand wird im Testmanual von 1966 weder berücksichtigt noch auch nur erörtert. Die Frage, ob eine Auswertung der TTCT nach diesen 4 Kriterien sinnvoll ist, soll bei der Besprechung von Untersuchungsergebnissen zu diesem Problem noch einmal aufgegriffen werden.

---

1) Crockenberg 1971, zitiert nach Crockenberg 1972; Harvey, Hoffmeister, Coates and White 197o; Mackler 1962, zitiert nach Dentler and Mackler 1964; Mackler and Shontz 1967; Mackler and Spotts 1965, zitiert nach Seiffge-Krenke 1974; Seiffge-Krenke 1974.

2) Zur Begründung s.S. 19

3) S. dazu Torrance (1964a)zu den figuralen Tests: "The ten-minute time limit is more than most children will use." (S. 127)

4) Cicirelli 1964, Long and Henderson 1964, beide zitiert nach Crockenberg 1972. Einige Informationen zu den beiden Stichproben gibt Torrance 1966.

2.223 Validität

Für die Besprechung der Validitätsüberprüfungen der TTCT wird die von der American Psychological Association vorgeschlagene Klassifikation der Validitätsarten in inhaltliche Validität, Übereinstimmungs- und Vorhersagevalidität und Konstruktvalidität übernommen[1]. Torrance (1966, 1967a) folgt in seiner Darstellung der Validitätsuntersuchungen ebenfalls dieser Klassifikation.

2.2231 Inhaltsvalidität

Das Konzept der Inhaltsvalidität geht davon aus, daß das durch den Test geforderte Verhalten eine Stichprobe des zu messenden Verhaltens außerhalb der Testsituation ist.
Der Validitätsnachweis beschränkt sich dabei auf die Feststellung der Repräsentativität der durch den Test gewonnenen Verhaltensstichprobe für das interessierende Alltagsverhalten. Diese Art von Validität beansprucht Torrance (1966, 1967a) für seine Tests. Er schreibt in dem Testhandbuch unter der Überschrift "Content Validity" :"The author does not believe that anyone can now specify the number and range of test tasks necessary to give a complete or even adequate assessment of a person's potentialities for creative behavior. He does believe that the sets of tasks assembled in the Figural and Verbal Batteries, Forms A and B, sample a rather wide range of abilities in such a universe... The test tasks selected for inclusion in the Figural and Verbal Forms A and B were chosen deliberately because it was believed that they call into play different parts of a universe of abilities that may legitimately be conceptualized as "creative thinking abilities". "(S. 23/24)
Die Annahme, daß das Testverhalten eine Stichprobe realen kreativen Verhaltens sei, erscheint auch unter Ausklammerung der Repräsentativitätsfrage nicht haltbar. Inhaltsvalidität kann bei Kenntnis- und Fertigkeitstests wie z.B. Rechtschreibe- oder Stenographietests beansprucht werden, nicht jedoch bei einem so komplexen Verhalten wie kreativer Produktion. Allein die starke motivationale Beeinflußbarkeit kreativen Verhaltens dürfte zu grundlegenden Unterschieden zwischen der Testsituation, die zur Bearbeitung des Tests zwingt, und der Alltagssituation, die fast immer mehrere Handlungsalternativen bietet, führen.

---

1) S. dazu Michel 1964

Krause (1972) lehnt die Möglichkeit, für Kreativitätstests inhaltliche Validität zu beanspruchen, ebenfalls ab: "Demgegenüber neigen wir eher zu der Ansicht, daß die Tests selbst nie Kreativität per se messen... In den gegenwärtig überblickbaren Tests werden nur Dispositionen für spätere kreative Leistungen gemessen." (S. 52) Nach diesen Überlegungen erscheint es nicht möglich, den Kreativitätstest von Torrance inhaltliche Validität zuzuerkennen.

2.2232 Übereinstimmungs- und Vorhersagevalidität

Die Übereinstimmungs- und die Vorhersagevalidität werden aufgrund der Kovariation des Test- und des zu messenden Kriteriumsverhaltens überprüft, wobei die Messung des Kriteriumsverhaltens Schwierigkeiten bereiten kann. Dies gilt mit Sicherheit für den Bereich der Kreativitätsforschung und in besonderem Ausmaß für Validitätsuntersuchungen mit Kindern, da Kinder nur äußerst selten Leistungen erbringen, deren Kreativität nicht angezweifelt werden kann. Wenn kreative Kinder mit Hilfe solcher Leistungen identifiziert werden könnten, bestünde die Validitätsprüfung beispielsweise darin, die Testergebnisse einer Gruppe kreativer Kinder mit denjenigen von Kindern, die sich nicht hervorgetan haben, zu vergleichen. Untersuchungen dieser Art wurden mit Erwachsenen häufig und erfolgreich durchgeführt[1], bei Kindern sind sie kaum möglich, weil deren kreative Leistungen wegen ihrer geringen sozialen Relevanz meist unerkannt oder jedenfalls unbeachtet bleiben.[2]

Ein zweiter Zugang zur Untersuchung der Übereinstimmungs- und Vorhersagevalidität bestünde darin, für eine Stichprobe von Kindern eine kreativitätsfordernde Situation zu schaffen und ihr Verhalten in dieser Situation von unabhängigen Beobachtern beurteilen zu lassen; anschließend könnten dann die Testergebnisse dieser Kinder mit ihren Verhaltensratings korreliert werden. Abgesehen von dem zeitlichen und personellen Aufwand besteht bei diesem Verfahren das Problem, ein für kreatives Verhalten repräsentatives Kriteriumsverhalten zu finden. Nur in diesen Schwierigkeiten kann die Begründung dafür gesehen werden, daß die überwiegende Zahl der von Torrance und seinen Mitarbeitern durchgeführten und der von Torrance (1966, 1967a) referierten Untersuchungen zur Übereinstimmungsvalidität Schulleistungen

---

1) Z.B. von Barron 1969, Cattell 1959, Cattel and Drevdahl 1955, MacKinnon 1962 und 1967

2) Wettbewerbe wie z.B. "Jugend forscht" o.ä., die die Identifikation einer "kreativen" Extremgruppe ermöglichen würden, werden meist nur für ältere Kinder und Jugendliche veranstaltet.

- erfaßt durch Schulleistungstests oder Zensuren - als Außenkriterium verwenden. In fast allen anderen Untersuchungen spielen Schulleistungen ebenfalls eine Rolle, weil hier Schätzurteile von Lehrern und Mitschülern über die Kreativität der Schüler das Kriterium bilden.

Die Untersuchungen, in denen Schulleistungen als Außenkriterium verwendet werden, sind zu zahlreich und verschiedenartig, um hier dargestellt zu werden. Torrance (1967a) stellt in einer Tabelle die Produkt-Moment-Korrelationen zwischen Kreativitäts- und Schulleistungstestergebnissen zusammen; die 65 Koeffizienten streuen von -.2o bis über .7o, der Median beträgt .28. Etwas niedriger sind die Korrelationen zwischen Kreativitätstestscores und Zensuren: die 114 Koeffizienten liegen zwischen -.3o und .69, Md=.21.

Die Interpretation der überwiegend positiven und häufig signifikanten, aber meist niedrigen Korrelationen bereitet einige Schwierigkeiten, vor allem wenn keine Informationen über die Art des Leistungsmaßes und über den vorangegangenen Unterricht, also Lernziele, Lehrmethoden, Unterrichtsstil etc., vorliegen. Torrance (1966) ist der Ansicht, daß aus vielen Gründen keine enge Beziehung zwischen Kreativität und Schulleistung erwartet werden kann und daß das widerspruchsvolle und verwirrende Bild, das die Untersuchungsergebnisse bieten, durch ein übervereinfachendes Vorgehen - vor allem durch mangelnde Berücksichtigung der genannten meßtechnischen und schulischen Variablen - zustandegekommen sei. So könne beispielsweise die Tatsache, daß bei einem Teil der Untersuchungen das gleiche Schulleistungskriterium durch traditionelle Intelligenztestmaße besser vorhergesagt werde als durch Kreativitätstestergebnisse, während es bei anderen Untersuchungen umgekehrt sei, dadurch erklärt werden, daß im ersten Fall eine stark lehrergeleitete Unterrichtsmethode mit rezeptivem Schülerverhalten, im zweiten Fall "kreative" Unterrichtsverfahren wie entdeckendes Lernen, Experimentieren etc. verwendet worden seien. Ebenso können Persönlichkeitsvariablen des Lehrers zur Erklärung inkonsistenter Untersuchungsergebnisse herangezogen werden.

Während diese Argumente durchaus plausibel erscheinen, fehlt eine grundsätzliche Erörterung der Beziehungen zwischen Kreativität und Intelligenz in ihrem Einfluß auf Schulleistungsvariablen. Es wird zwar eine Reihe von Untersuchungsergebnissen dargestellt, aber die

Bedeutung der Kreativität für die Schulleistung wird überschätzt[1] und widersprüchliche Ergebnisse bleiben unerörtert[2].
Die Ergebnisse dreier von Torrance nicht berücksichtigter Untersuchungen von Dewing (197o b), Yamamoto und Chimbidis (1966) und von Yamamoto (1967a) zwingen zu einer abweichenden Beurteilung des Zusammenhangs zwischen Kreativität und Schulleistung. Dewing (197o b) untersuchte in 8 Schulen an insgesamt 32o Schülern siebter Klassen den Zusammenhang zwischen einem auf 4 Aufgaben beruhenden Gesamtwert (Auswertung nach Flüssigkeit und Originalität) und dem Notendurchschnitt des Halbjahreszeugnisses. Die Koeffizienten der 8 Schulen lagen zwischen -.19 und .28, Md=.o5; die Korrelation für alle 32o Schüler war .o7. Yamamoto (1967a) überprüfte eine Annahme von Torrance, daß das Phänomen der Overachiever durch ihre im Vergleich zu Underachievern höhere Kreativität erklärt werden könne und kam zu einem negativen Ergebnis. Er schließt aus seinen Ergebnissen, daß "creativity does not contribute to school achievement, as measured on some of the current achievement batteries, to any significant degree" (S. 324). In der Arbeit von Yamamoto und Chimbidis wurden die korrelativen Zusammenhänge zwischen Kreativität (1), Intelligenz (2) und Schulleistung (3) bei 79o Fünftkläßlern untersucht; die multiple Korrelation für $R_{3.12}$ war .57, die einfache Korrelation zwischen Intelligenz und Schulleistung betrug .56.
Während viele Untersuchungen von Torrance und seinen Mitarbeitern an Stichproben von hochbegabten Schülern mit eingeschränkter Streuung der Intelligenztestwerte durchgeführt wurden, basieren die Ergebnisse der Untersuchung von Yamamoto und Chimbidis auf einer sehr

---

1) So werden Untersuchungsergebnisse von Bish (1964) und Cicirelli (1965) als ermutigend hinsichtlich der Übereinstimmungsvalidität bezeichnet, obwohl Cicirellis Validitätskoeffizienten nach Auspartialisierung der Intelligenz zwischen .o4 und .3o betragen (Md=.14)! Vgl. Torrance 1966, S. 48.

2) Torrance (1966) berichtet z.B. kommentarlos, daß Cicirelli nach Auspartialisierung der Intelligenz niedrigere Korrelationen zwischen Kreativitätsmaßen und Schulleistungstestergebnissen erhielt und daß bei Bish (1964) die Korrelationskoeffizienten nach Auspartialisierung der Intelligenz angestiegen seien. (Vgl. Torrance 1966, S. 47-48). Torrance (1967a) stellt in bezug auf die Ergebnisse von Bowers (1959) fest, daß sowohl Intelligenz als auch Kreativität zur Schulleistung beitragen, letztere jedoch mehr auf niedrigerem Intelligenzniveau, während er aufgrund seiner Untersuchungsergebnisse zu dem Schluß kommt, daß Intelligenz und Kreativität auf der unteren Hälfte oder dem unteren Viertel des Intelligenzkontinuums enger zusammenhängen als auf der oberen Hälfte oder dem oberen Viertel. (Vgl. Torrance 1967, S. 147 und S. 15o)

großen und in den Intelligenztestwerten relativ weit streuenden Stichprobe (δ =13,6). Somit muß der Versuch von Torrance, mit Hilfe des Außenkriteriums "Schulleistungen" die Übereinstimmungsvalidität seines Tests nachzuweisen,als gescheitert betrachtet werden. Ob allerdings ein Erfolg dieser Bemühungen wirklich als Validitätsnachweis gewertet werden könnte, erscheint mehr als zweifelhaft, da sowohl in Schulleistungstests als auch in den traditionellen Klassenarbeiten überwiegend Reproduktion von Gelerntem verlangt wird; die Validität dieses Validitätskriteriums ist zumindest als ungesichert zu betrachten.

Die 2. Gruppe von Untersuchungen zur Übereinstimmungsvalidität verwendete Schätzurteile als Kriterium. Bevorzugte Methode bei den Lehrerratings war,die Lehrer nach den Schülern mit der höchsten Flüssigkeit der Ideenproduktion, nach den flexibelsten und originellsten Schülern ihrer Klasse und nach den Schülern, die ihre Einfälle am besten elaborieren, zu fragen. Z.T. wurde auch nach denjenigen Schülern gefragt, die in diesen Merkmalen am schlechtesten abschneiden. Zu den 4 Kriterien, nach denen gefragt wurde, erhielten die Lehrer eine kurze Erläuterung (1-4 kurze Sätze). Die Anzahl der zu nennenden Schüler wurde teilweise freigestellt und z.T. auf 5 begrenzt. Die Testergebnisse der beiden Extremgruppen bzw. der nominierten und nicht nominierten Schüler wurden dann auf Mittelwertsunterschiede in den jeweils entsprechenden Kreativitätsmaßen untersucht. Von den 7 Untersuchungen in 4.-6. Klassen, zu denen Torrance (1966) einige Angaben macht, erzielten 6 signifikante Mittelwertsunterschiede für alle oder mehrere der 4 Kreativitätsmaße (die nur bei einer Untersuchung in verbale und figurale Maße aufgegliedert waren); die Flüssigkeitsvariable scheint von den Lehrern am sichersten beurteilt werden zu können. In keiner der Untersuchungen wurde die mögliche Einflußvariable Schulleistung und in nur wenigen der Einfluß der Intelligenz überprüft oder gar kontrolliert.
Seiffge-Krenke (1974) übersetzte den von Torrance (1966) vorgeschlagenen Fragebogen für die Lehrerratings ins Deutsche und ließ 6 Lehrer vierter Klassen ihre 5 besten und 5 schlechtesten Schüler hinsichtlich der 4 Kreativitätsmaße nominieren. Anschließend wurde untersucht, wie gut die nominierten Schüler aufgrund der 4 Kreativitätsgesamtscores (hier wurden die verbalen und figuralen Kreativitäts-

maße summiert) und des Rohwertes der Kurzform der Standard Progressive Matrices von Raven (1960) mit Hilfe einer schrittweisen Diskriminanzanalyse den insgesamt 8 Extremgruppen zugeordnet werden können und welcher der beiden Prädiktoren - das Kreativitätsmaß der TTCT oder der Intelligenztestwert - das Kriterium der Gruppenzugehörigkeit besser vorhersagt. In jedem Fall sagte das Kreativitätsmaß die Zugehörigkeit zu einer der beiden entsprechenden Kriteriengruppen besser vorher als der Intelligenztestwert, dessen Hinzunahme die Zahl der Fehlklassifikationen um 1-3 Schüler verringerte. Bei 4 Gruppen von 40 zu klassifizierenden Schülern lag die Zahl der mit Hilfe des Kreativitätsmaßes richtig zugeordneten Fälle zwischen 29 und 32. Seiffge-Krenke läßt bei ihrer sehr positiven Beurteilung dieses Ergebnisses völlig außer acht, daß das immer noch relativ selten eingesetzte statistische Verfahren der Diskriminanzanalyse bei der Verwendung von Extremgruppen nur zu einer völlig unverbindlichen Validitätsschätzung führen kann, weil keinerlei Erfahrungswerte dafür vorliegen, wie gut Angehörige von Extremgruppen durch Tests, deren Validität aufgrund anderer Methoden nachgewiesen wurde, klassifiziert werden können. Somit kann auch nicht beurteilt werden, ob der Anteil von etwa 25 % Fehlklassifikationen hoch oder niedrig ist. Es kann lediglich gesagt werden, daß das sehr spezifische Kriterium - d.h. die Zugehörigkeit zur Gruppe der besonders "Flüssigen", Flexiblen etc. - durch einen ebenfalls spezifischen Prädikator - den Testgesamtwert für Flüssigkeit, Flexibilität etc. - besser vorhergesagt wird als durch das Ergebnis der "Progressive Matrices", eines Intelligenztests also, der als gutes Maß der allgemeinen Intelligenz gilt, aber mit dem von den Lehrern beurteilten Kriteriumsverhalten im Vergleich zu anderen Intelligenztests relativ wenig gemeinsam haben dürfte.[1] Darüber hinaus wurde nicht berücksichtigt, daß das Lehrerurteil durch die Schulleistungen beeinflußt

---

1) Seiffge-Krenke macht weder Angaben über den Zusammenhang zwischen Intelligenztestwerten und Gruppenzugehörigkeit noch über die Korrelation zwischen Intelligenz und den 4 Kreativitätsmaßen, abgesehen von der Beziehung zwischen Intelligenz und Originalitätsscore, die mit .53 als einzige signifikant korrelieren. Grote, Hajek u.a. (1969) und Krause (1972) fanden dagegen signifikante bzw. hochsignifikante Beziehungen zwischen einem Gesamtwert für den Kreativitätstest von Wallach und Kogan, bei dessen Konstruktion Unabhängigkeit von Intelligenztestergebnissen eines der beiden Hauptziele war, und dem Ergebnis des Frankfurter Analogietests (bei Grote, Hajek u.a.), der vor allem Thurstones Reasoning-Faktor mißt, bzw. dem Gesamtwert des Leistungs-Prüf-Systems von Horn (bei Krause).

worden sein kann. Eventuell wäre mit Hilfe dreier Zensuren eine ebenso gute Trennung der Gruppen möglich gewesen wie mit Hilfe der beiden von Seiffge-Krenke verwendeten Prädiktoren. Wegen der zur Beurteilung von Untersuchungsergebnissen notwendigen Vergleichsdaten müßten in Untersuchungen, die neue Verfahren oder alte Verfahren auf eine neue Art einsetzen, zusätzlich traditionelle Methoden verwendet werden. Für die Untersuchung von Seiffge-Krenke könnte das z.B. bedeuten, daß zusätzlich zu der Diskriminanzanalyse eine multiple Regressionsanalyse mit Kreativitäts- und Intelligenztestergebnissen sowie möglichst auch mit Schulleistungsvariablen als Prädiktoren und dem Lehrerrating als Kriterium gerechnet oder zumindest eine einfache Korrelationstabelle mit den genannten Variablen wiedergegeben wird.[1)]

Während in den von Torrance (1966) und in den von Yamamoto (1967b) in einem Sammelreferat erwähnten Untersuchungen, die Lehrerratings als Validitätskriterium verwendeten, nur Mittelwertsvergleiche oder Varianzanalysen - meist mit signifikanten Ergebnissen - durchgeführt wurden, gibt Torrance (1967a)in einer Tabelle die Korrelationskoeffizienten von 4o Untersuchungen zur Validität der TTCT mit Lehrerschätzurteilen als Kriterium wieder: sie streuen von -.19 bis über .7o, der Median ist .2o. Da keinerlei zusätzliche Informationen gegeben werden, nicht einmal über die Zahl der Versuchspersonen, kann nicht beurteilt werden, ob die höheren oder weniger hohen

---

1) Statt dessen werden durch multiple Regressionsanalysen 28 Kreativitätsergebnisse durch die Prädiktoren Intelligenz, Lehrerstil und Alter vorhergesagt! Da es sich um eine Stichprobe von Viertkläßlern handelt, erhebt sich die Frage, welche Hypothese hinter der Verwendung des Alters als Prädiktorvariable steht; die Streuung des Alters in einer 4. Klasse ist so gering, daß keine Korrelation mit den Kreativitätstestergebnissen erwartet werden kann, hinzu kommt, daß die ältesten Kinder Repetenten und später Eingeschulte sind, die vermutlich eher schlechtere Testergebnisse erzielen als ihre Klassenkameraden! Auch die Aufforderung, die Interkorrelationen der 3 Prädiktoren in einer Tabelle zu betrachten, erscheint überflüssig; die Intelligenztestergebnisse sind alterskorrigiert, der Lehrerstil von Lehrern vierter Klassen kann ebenfalls nicht mit dem Alter der Schüler korrelieren und wieso in unausgelesenen Grundschulklassen der Lehrerstil mit der Intelligenz kovariieren sollte, bleibt auch unklar.

Koeffizienten die zuverlässigeren Schätzungen des Zusammenhangs darstellen. Der Median der Koeffizienten deutet allerdings darauf hin, daß die Korrelationen so niedrig sind, daß sie beispielsweise allein durch die Beziehung sowohl der TTCT als auch der Lehrerratings zur Intelligenz zustandegekommen sein könnten.
Spätere Untersuchungen erbrachten ebenfalls keine eindeutigen Validitätshinweise. Haddon und Lytton (1971) ließen Englisch-Lehrer die Kreativität ihrer Schüler auf einer 5stufigen Skala einschätzen und erhielten eine Korrelation von .286 zwischen den Ratings und dem TTCT-Gesamtwert von 148 Schülern.
Dewing (197oa) ließ sowohl Lehrer als auch Schüler die 2o Prozent der Schüler nominieren, die sich in ihren Aktivitäten in der Schule am kreativsten zeigten, und setzte die Resultate zu den Scores zweier Fragebogen - der Torrance Creative Leisure Interests Check-list und der Golann Creative Motivation Scale -, zu den Ratings eines Phantasie-Aufsatzes durch einen über die übrigen Ergebnisse nicht informierten Beurteiler und zu den TTCT-Ergebnissen in Beziehung. Lehrer- und Schülerratings korrelierten untereinander hoch und mäßig bis insignifikant mit den Aufsatzbeurteilungen. Die Ergebnisse der beiden Fragebogen korrelieren niedrig, aber in den meisten Fällen signifikant positiv miteinander, für verschiedene Schulen völlig unterschiedlich mit den Schülernominierungen und insignifikant bis niedrig positiv mit den Aufsatzbeurteilungen. Leider waren keine Produkt-Moment- oder Rangkorrelationen der 5 Verfahren mit den TTCT-Ergebnissen mitgeteilt, sondern die 2o Prozent der Schüler mit den höchsten Kreativitätstestergebnissen wurden mit einer Kontrollgruppe nach Geschlecht, IQ und Schule gematcht und dann in eine von 3 "Performanz"-Gruppen eingeteilt (in die untere Gruppe, wenn sie nach keinem der 5 Verfahren zu den oberen 2o Prozent gehörten, in die mittlere, wenn sie nach einem der 5 Maße, und zur oberen Gruppe, wenn sie nach 2, 3, 4 oder 5 Maßen zu den oberen 2o Prozent zählten). Die Anzahlen der kreativen und der Kontrollgruppen-Schüler unterschieden sich in den 3 Gruppen hochsignifikant; die tetrachorische Korrelation für die Beziehung zwischen dem dichotomisierten TTCT-Wert und der Zugehörigkeit zu einer der 3 Performanz-Gruppen wird mit .59, der Kontingenz-Koeffizient C mit .392 angegeben. An dieser relativ aufwendigen Untersuchung wird die Schwierigkeit, valide Außenkriterien für Kreativitätstests

zu finden, besonders deutlich. Wegen des geringen oder fehlenden Zusammenhangs zwischen den verschiedenen Kriterien sollten Untersuchungen, die nur ein Validitätskriterium verwenden, besonders vorsichtig interpretiert werden.

Holland (1959) führte eine Untersuchung über die Validität von Lehrerratings durch; zu beurteilen waren eine Reihe von Persönlichkeitsmerkmalen an über 1000 Hochbegabten. Obwohl es sich um ganz verschiedene Merkmale handelte, korrelierten die Ratings untereinander hoch, was als Auswirkung eines Halo-Effekts interpretiert werden kann. Holland wählte die Beurteilungsdimension "Reife", die mit allen anderen am höchsten korrelierte, und untersuchte, welche Persönlichkeitsmerkmale diejenigen Schüler, die von ihren Lehrern am höchsten eingestuft worden waren, nach dem Sixteen Personality Factor Questionnaire von Cattell und 2 anderen Fragebogen hatten. Es waren intelligente, anstrengungsbereite und gewissenhafte Schüler mit guten Schulleistungen und Führungseigenschaften. Holland folgert aus seinen Ergebnissen, daß Lehrerratings Schulleistungen und Führerqualitäten besser vorhersagen als kreative Leistungen.

Wallach (1970) berichtet von 4 kleineren Untersuchungen über die Fähigkeit von Lehrern, die Kreativität ihrer Schüler zu beurteilen, und kommt aufgrund der Ergebnisse zu dem Schluß:"the teacher ratings predict the creativity indices to the extent that the ratings are judgments of general intelligence and the creativity indices share variance with intelligence." (S. 1230)

Wasik (1974) untersuchte, welche Variablen die Beurteilung kreativer Problemlösungsfähigkeiten durch Lehrer determinieren. 147 Zehntkläßler wurden von 4 unabhängig arbeitenden Lehrerteams im Hinblick auf diese Fähigkeiten,die so beschrieben waren, daß die Rater von beobachtbarem Verhalten ausgehen konnten, beurteilt. Die Schüler wurden mit einer Batterie semantischer Guilford-Tests (5 Aufgaben zur Erfassung der divergenten, eine zur Erfassung der konvergenten Produktion, je eine Aufgabe zur Erfassung der Fähigkeiten des Erkennens und Bewertens semantischer Implikationen) untersucht. Die Testergebnisse wurden durch eine multiple Regressionsanalyse - die Judgment Analysis Technique von Christal (1961) - zu den 4 Ratings in Beziehung gesetzt. Es zeigte sich, daß die 4 Lehrerteams - obwohl

Lehrer verschiedener Fachrichtungen - das gleiche Bezugssystem bei der Beurteilung der Schüler verwendeten und daß von den 8 S-I-Fähigkeiten nur 3 für die Ratings bedeutsam waren (eine Fähigkeit der divergenten, eine der konvergenten Produktion und der Evaluationsfaktor). Durch Hinzunahme des IQ vom California Test of Mental Maturity konnte die Vorhersage der Ratings erheblich verbessert werden.
Somit kann zusammenfassend gesagt werden, daß die Validitätsuntersuchungen, die Lehrerratings als Außenkriterium verwendeten, keine sichere Beurteilungsgrundlage für die Validität der TTCT bieten, und zwar vor allem wegen der zweifelhaften Validität des Kriteriums und wegen des geringen Zusammenhangs zwischen Testergebnis und Kriterium.

In einer weiteren Gruppe von Untersuchungen wurden die Schätzurteile der Mitschüler als Validitätskriterium verwendet. Torrance (1966) berichtet von einer nicht näher bezeichneten Untersuchung, in der er Schüler der Klassen 1 bis 6 nach den Klassenkameraden fragte, die die meisten guten Ideen haben, die gleich einen neuen Einfall haben, wenn ein Vorhaben fehlschlägt, die am meisten reden, die viele gute Ideen haben, aber niemand etwas davon erzählen u.ä. Mit diesen inhaltlich sehr heterogenen Fragen und einigen aus dem Ask-and-Guess-Test abgeleiteten Maßen führte er eine Faktorenanalyse durch, um mit Hilfe der Ergebnisse eine geeignete Kriteriumsfrage zu identifizieren (es war die Frage nach den meisten guten Ideen). Die Testergebnisse der Schüler, die zu diesem Kriterium 5 oder mehr Nominierungen erhielten, wurden dann mit denjenigen der nichtgenannten Schüler verglichen. Daß erstere signifikant höhere Ergebnisse in den TTCT erzielten als letztere, wird von Torrance als Validitätsnachweis gewertet.
Yamamoto (1967b) berichtet von einer Untersuchung mit ähnlicher Fragestellung (Torrance 1959a), in der nach den Ergebnissen der TTCT hochkreative Kinder für 5 Kreativitätskriterien (in der Art der oben erwähnten Fragen) nicht mehr Nominierungen erhielten als die hochintelligenten, aber weniger kreativen Schüler.
Yamamoto (1964a) führte eine Untersuchung mit 428 Schülern der Klassen 7 bis 12 durch, in denen er die Schüler klassenweise nach den Ergebnissen der TTCT in 3 Gruppen einteilte und sie nach den einfallsreichsten, originellsten etc. Mitschülern fragte. (Die Begriffe waren je-

weils ganz kurz erläutert.) Insgesamt stimmten die Ergebnisse der Schülerratings mit den Gruppierungen aufgrund der Testergebnisse in der 8. und 9. Klasse recht gut überein, in den letzten Klassen ("senior-high level") stimmten nur noch die Ratings der Jungen mit den TTCT-Scores überein. Für die einzelnen Kreativitätsmaße fielen die Ergebnisse z.T. unterschiedlich aus. Für die Gesamtgruppe aus allen Klassen ergab sich eine Korrelation von .24 zwischen Flüssigkeits-Testscores und den Schülerratings für Flüssigkeit und von .22 für die Beziehung zwischen Flexibilitäts-Testwert und Flexibilitäts-Ratings.[1] Angaben über die Korrelation der ebenfalls erhobenen Intelligenztestwerte sowohl mit den Ergebnissen der TTCT als auch mit den Schülerratings werden nicht gemacht. So besteht auch bei dieser Untersuchung die Möglichkeit, daß die zwar signifikanten, aber niedrigen Zusammenhänge zwischen Testergebnissen und Ratings durch die Korrelation beider Variablen mit der Intelligenz der Schüler zu erklären sind.
Seiffge-Krenke (1974) führte in den 6 vierten Klassen nicht nur die bereits erwähnte Untersuchung zu der Frage durch, wie gut die TTCT-Ergebnisse eine aufgrund von Lehrerurteilen vorgenommene Klassifikation von Schülern reproduzieren können (vgl. S.36 -38), sondern sie untersuchte dieselbe Frage auch für eine aufgrund von Schülerratings vorgenommene Klassifikation. Sie verwendete dieselben Prädiktoren, dasselbe statistische Verfahren und erhielt vergleichbare Ergebnisse (s. S. 37). Welche Schlußfolgerungen hinsichtlich der Validität der TTCT daraus abgeleitet werden können, kann nicht beurteilt werden (s. dazu S. 37 - 38).
Insgesamt können auch die Validitätsuntersuchungen, die Schätzurteile von Schülern als Außenkriterium verwenden, nicht sehr günstig beurteilt werden. Keine der besprochenen Arbeiten erbrachte Ergebnisse, die als eindeutiger Nachweis der Validität der TTCT gewertet werden könnten.

Bisher wurden nur relativ wenige Untersuchungen zur Vorhersagevalidität durchgeführt.
Torrance (1966) berichtet von einer einzigen Untersuchung - von

---

1) Die beiden Korrelationskoeffizienten zu dieser Untersuchung werden in Yamamoto (1967b)mitgeteilt.

Erickson (1966) -, in der alle 1959 mit den TTCT untersuchten High School Seniors der Universität von Minnesota nach 6 Jahren über kreative Leistungen und Aktivitäten befragt wurden. Die Checklist kreativer Betätigungen enthielt Fragen nach allen möglichen herausragenden Leistungen in Beruf oder Studium, aber auch nach Religionswechsel, dem Erlernen einer Sprache, nach Zeitschriftenabonnements (!) u.ä.. Aufgrund von 44 eingegangenen Beantwortungen wurden die Korrelationen zwischen den Gesamtwerten der 4 Kreativitätsmaße und einem aus den Befragungsergebnissen gebildeten Kriteriumsmaß berechnet; sie lagen bei .27 für Flüssigkeit, .24 für Flexibilität, .17 für Originalität und .16 für Elaboration. Angesichts der geringen Retest-Reliabilitäten der TTCT (vgl. S. 26 - 28) sind diese niedrigen Koeffizienten nicht verwunderlich, erstaunlich ist dagegen ihre Interpretation durch Torrance (1966 und 1967a)als Hinweis auf die prädiktive Validität seines Verfahrens. Torrance (1972c)berichtet von einer Untersuchung zur Vorhersagegültigkeit der TTCT über 12 Jahre. Die Tests waren mit der gesamten High School-Population der Minnesota University von 1959 durchgeführt worden; 236 der 392 Probanden beantworteten 1971 einen Fragebogen über verschiedene biographische Daten, über kreative Betätigungen und Aktivitäten nach der High school und ihre Wünsche und Pläne für die Zukunft. Die daraus entwickelten 3 Kriteriumsmaße - Quantität, Qualität kreativer Leistungen und Zukunftspläne - wurden mit den 4 Gesamtwerten mehrerer verbaler und einer figuralen Aufgabe korreliert. Die Koeffizienten für die unteren beiden Klassen streuten von .14 bis .41, der Median war .235; nur einer der 6 Koeffizienten war signifikant. Für höhere Klassenstufen ergaben sich allerdings z.T. wesentlich engere Zusammenhänge. Die prädiktive Validität der TTCT war für Männer erheblich höher als für Frauen.

Haddon und Lytton (1971) versuchten, die durch Tests ermittelten Schulleistungen in der Muttersprache und in Mathematik durch verschiedene, 4 Jahre zuvor erhobene Testergebnisse - den Verbal Reasoning Quotient und den verbalen und figuralen Gesamtwert sowie den aus beiden ermittelten totalen Gesamtwert von 5 TTCT-Aufgaben - vorherzusagen. Die Koeffizienten für 97 bzw. 4o 11-12jährige Schüler zweier Schulen waren für Englisch

r=.2o und .o3 bei der Vorhersage durch den figuralen,

r=.46 und .18 bei der Vorhersage durch den verbalen,
r=.38 und .11 bei der Vorhersage durch den totalen Gesamtwert.

Bei der Vorhersage der Mathematikleistung ergaben sich folgende Korrelationen:

r =.09 und .04 für den figuralen Gesamtwert,
r =.39 und -.05 für den verbalen Gesamtwert,
r =.27 und .01 für den totalen Gesamtwert als Prädiktor.

1972 führte Cropley eine Untersuchung zur prädiktiven Validität von Kreativitätstestitems durch, in der neben 4 Guilford-Aufgaben und einem Intelligenztestwert auch 2 Aufgaben von Torrance - Tin Can Uses und Circles - verwendet wurden. Versuchspersonen waren 111 Schüler 7. High School-Klassen, vorhergesagt wurden extracurriculare Aktivitäten auf künstlerischem Gebiet über einen Zeitraum von 5 Jahren. Aus Befragungsergebnissen der Schüler wurden 4 Kriteriumsmaße gebildet und u.a. ihre Produkt-Moment-Korrelationen mit den Prädiktorvariablen berechnet. Die für Jungen und Mädchen getrennt ermittelten Koeffizienten liegen zwischen -.20 und +.27, der Median ist .034. Für beide Geschlechter wurde nur je ein auf dem 5 Prozent-Niveau signifikanter Koeffizient von .27 ermittelt und zwar für den gleichen Prädiktor - Tin Can Uses-, aber für verschiedene Kriterien (literarische Aktivitäten bei Jungen, Betätigungen auf dem Gebiet der Musik bei Mädchen). Auch abgesehen von möglichen Einwänden gegen die subjektiven Kriteriumsmaße kann dieses Ergebnis keineswegs als Nachweis der prädiktiven Validität der beiden Torrance-Aufgaben gewertet werden. Somit liegen bisher für jüngere Schüler noch keinerlei Belege für die prädiktive Validität der TTCT vor. Holland (1968) kommt in seiner insgesamt sehr positiven Beurteilung des Verfahrens zu dem Fazit: "Unless tests like the TTCT are tied to reality by studies of their external validity, it will never be known,whether high scores identify original people or crazy bricklayers and packers. The current evidence about the TTCT implies that it may have some predictive validity, but explicit evidence is needed about its ability to forecast socially relevant creative behavior." (S. 298)

2.2233 Konstruktvalidität

Das Konzept der Konstruktvalidität ist für alle Tests, die ein nicht operational definiertes Merkmal messen, von Bedeutung[1] und

---

1) Vgl. Cronbach and Meehl 1955

spielt somit bei der Validierung von Kreativitätstests, für die weder angenommen werden kann, daß sie eine Stichprobe realen kreativen Verhaltens erheben, noch daß unmittelbar zugängliche und vor allem valide Außenkriterien zur Verfügung stehen, eine zentrale Rolle. Das zu messende Konstrukt wird in ein theoretisches Bezugssystem eingeordnet, aus dem empirisch überprüfbare Hypothesen abgeleitet werden[1]; die Ergebnisse solcher Überprüfungen können zu Modifikationen sowohl des Meßinstruments als auch des Konstrukts selbst führen.[2]

Das Verfahren der Konstruktvalidierung erfordert somit ein theoretisches Bezugssystem, eine Reihe empirisch überprüfbarer Hypothesen, die aus diesem System abzuleiten sind, und eine möglichst große Methodenvielfalt bei der Hypothesenprüfung[3].
Es soll nun überprüft werden, ob mit Hilfe der Untersuchungen zur Konstruktvalidität der TTCT ein überzeugenderer Validitätsnachweis erbracht werden konnte als mit den Überlegungen zur Inhaltsvalidität und den Versuchen, die Übereinstimmungs- und die Vorhersagevalidität des Verfahrens zu belegen.
Zunächst wird untersucht, ob die Abgrenzung der TTCT gegen Verfahren zur Messung der allgemeinen Intelligenz gelungen ist, die nach Torrances theoretischen Vorannahmen[4] im Verlauf einer Konstruktvalidierung erfolgen müßte. Diese Abgrenzung könnte mit Hilfe der Faktorenanalyse demonstriert werden oder dadurch, daß die Korrelationen der Kreativitätstestitems untereinander mit denjenigen zwischen Intelligenz-Subtests und den Kreativitätsaufgaben verglichen werden[5]; wenn erstere deutlich höher sind als letztere, kann angenommen werden, daß der Kreativitätstest ein von der Intelligenz unterscheidbares Konstrukt mißt.

---

1) Vgl. Michel 1964

2) S. dazu Lienert 1967²

3) Vgl. Cronbach and Meehl 1955, Michel 1964, Lienert 1967²

4) S. dazu S. 2o - 22

5) Intelligenzsubtest-Ergebnisse oder Intelligenztest-Gesamtwerte werden zu Kreativitätsitem-Scores in Beziehung gesetzt, da die TTCT überwiegend Einzelaufgaben und nur 2 Gruppen aus formal gleichen Items enthalten (diese werden wie Subtests behandelt). Die Gleichsetzung von Subtest- und Kreativitätsitem-Ergebnissen dürfte insofern statistisch unbedenklich sein, als auch letztere zu mehr als einer richtigen Antwort führen können. Intelligenztest-Gesamtwerte sollten dagegen nicht mit Kreativitätsitem-Scores korreliert werden, da erstere als zusammengesetzte Maße mehr Varianzquellen beinhalten und daher eine erhöhte Wahrscheinlichkeit besteht, daß Itemscores höher mit ihnen als untereinander korrelieren (vgl. Guilford 1968, zitiert nach Bennett 1973).

Wallach und Kogan (1965) vermuteten bereits aufgrund der zum Zeitpunkt ihrer Untersuchung vorliegenden Ergebnisse, daß die TTCT diese Forderung nicht erfüllen, und Wallach (1970[3]) bestätigte diese Vermutung aufgrund neuerer und methodisch sorgfältigerer Untersuchungen.

Daß die interne Konsistenz der Items der TTCT relativ niedrig ist, wurde bereits belegt (vgl. S.29 - 30); es wurde allerdings auch darauf hingewiesen, daß dieser Punkt Torrance wegen seiner Konzeption des Testverhaltens als einer Stichprobe heterogener kreativer Verhaltensweisen wenig Probleme bereitet.
Schwierig, wenn nicht unmöglich wird die im Rahmen einer Konstruktvalidierung erforderliche Abgrenzung jedoch dadurch, daß die TTCT-Ergebnisse mit Intelligenztestwerten in den meisten Untersuchungen, die Stichproben mit normaler Verteilung der Intelligenztestwerte verwenden, mäßige bis mittlere positive Korrelationen aufweisen, Korrelation also, die z.T. ebenso hoch sind wie die Interkorrelationen der Kreativitätsmaße.
Die folgende Tabelle gibt einen Überblick über in diesem Zusammenhang interessierende Untersuchungsergebnisse (sie enthält keine Arbeiten, die Stichproben von hochintelligenten oder hochkreativen Kindern verwenden).

Tab. 1: Überblick über Untersuchungsergebnisse zur Beziehung der Kreativitätsmaße der TTCT untereinander und zur Beziehung zwischen Kreativitäts- und Intelligenztestergebnissen

Autor(en)	Vpn	Interkorrelationen der Kreativitätstests bzw. -maße	Kreativitäts-Intelligenz-Korrelation
Cicirelli 1965	641 Sechstkläßler	keine Angaben	2 verbale Kreativitätsmaße/IQ des California Test of Mental Maturity r_1=.24, r_2=.23 2 figurale Kreativitätsmaße/IQ r_1=.09, r_2=.20
Cropley 1966	320 Siebtkläßler	Tin Can Uses/Circles r = .220	Tin Can Uses/Lorge-Thorndike-IQ r=.194 (Verbal IQ) r=.231 (Performance IQ) Circles/IQ r=.145 (Verbal IQ) r=.171 (Performance IQ)
Dacey und Ripple 1969	219 Achtkläßler	keine Angaben	durchschnittliche Korrelation zwischen 4 aus 2 verschiedenen Aufgaben abgeleiteten Kreativitätsmaßen und IQ (ohne nähere Angabe) $\bar{r}$ =.19
Dewing 1970 b	394 Siebtkläßler	keine Angaben	Gesamtwert aus den verbalen Flüssigkeits- und Originalitätsscores für 4 Torranceaufgaben/A.C.E.R.-IQ r=.20; Gesamtwert aus den figuralen Flüssigkeits- und Originalitätsscores/ A.C.E.R.-IQ r=.13
Haddon und Lytton 1968	211 11-12-jährige Elementarschüler	keine Angaben	Gesamtwert aus verbalen und figuralen Kreativitätsmaßen (vor allem Flexibilität und Originalität) aus 5 Torrance-Aufgaben/Verbal Reasoning Quotient r=.480
Haddon und Lytton 1971	148 15jährige Sekundarschüler (aus der Stichprobe der Untersuchung von 1968)	die Kreativitätsgesamtwerte korrelierten über 4 Jahre hinweg: figural/figural r=.415 figural/verbal r=.438 figural/Total r=.495 verbal-figural r=.331 verbal/verbal r=.572 verbal/Total r=.550 Total/figural r=.448 Total/verbal r=.598 Total/Total r=.621	dasselbe Kreativitätsgesamtmaß wie bei Haddon und Lytton (1968)/VRQ, gemessen 4 Jahre früher: r=.615

Fortsetzung Tabelle 1

Autor(en)	Vpn	Interkorrelationen der Kreativitätstests bzw. -maße	Kreativitäts-Intelligenz-Korrelation
Hauck und Thomas 1972	8o 4.-, 5.- und 6.-Kläßler	keine Angaben	Kreativitätsgesamtwert über alle Maße/Lorge-Thorndike-IQ r=.29
Lytton und Cotton 1969	143 14jährige Sekundarschüler	verbaler und figuraler Gesamtwert r=.43	dasselbe Kreativitätsmaß wie bei Haddon und Lytton (1968)/VRQ r=.17 (p=.o5)
Madaus 1967	6o3 Neunt- und Zehntkläßler	alle 8 verbalen und figuralen Kreativitätsmaße .24 ≷ r ≷ .94; Md=.39	8 Kreativitätsmaße/SCAT-Total -.o9 ≷ r ≷.16, Md=.o65
O'Bryan und Mac Arthur 1969	85 neunjährige Jungen	Unusual Uses, Figure Completion, Picture Construction, Cirles -.o6 ≷ r ≷ .22, Md=.185	Kreativitätsmaße jeder Aufgabe/OTIS verbal .o5 ≷ r ≷ .35, Md=.21 Kreativitätsmaße jeder Aufgabe/OTIS nonverbal .o9 ≷ r ≷ .39, Md=.13
Wodtke 1964	über 5oo Schüler 2.-5. Klassen	Wodtke macht nur Angaben zur Retest-Reliabilität (s.S. 27)	4 verbale Kreativitätsmaße/Lorge-Thorndike-IQ .27 ≷ r ≷ .5o, Md=.43 4 figurale Maße/IQ .1o ≷ r ≷ .33, Md= .24
Yamamoto 1964b	272 Schüler 9.-12.Klassen	keine Angaben	Kreativitäts-Gesamtwert für 5 Torrance-Aufgaben/Lorge-Thorndike-IQ (verbal) r=.3o
Yamamoto und Chimbidis 1966	6o4 Fünftkläßler	keine Angaben	Kreativitäts-Gesamtwert für 4 Torrance-Aufgaben/Lorge-Thorndike-IQ (verbal) r=.26
Yamamoto 1965b	461 und 827 Fünftkläßler	keine Angaben	Kreativitäts-Gesamtwert für 5 Torrance-Aufgaben/Lorge-Thorndike-IQ (verbal) Korrelation ohne Korrektur: r=.33 für N=461 r=.39 für N=827 Korrelation nach Korrektur wegen Streuungseinschränkung r=.45 für N=461 r=.48 für N=827 Korrelation nach einer weiteren Korrektur wegen Unreliabilität der Kreativitätswerte: r=.51 für N=461 r=.54 für N=827

Abgesehen von den Ergebnissen von Madaus (1967) korrelieren die Kreativitätstestergebnisse signifikant und z.T. substantiell positiv mit den verschiedenen Intelligenztestwerten oder sie zeigen zumindest etwa ebenso enge Beziehungen zur Intelligenz wie zu anderen Maßen der TTCT.
Eine Erklärung für die abweichenden Resultate von Madaus kann nicht gegeben werden; seine Stichprobe umfaßte alle 9. und 1o. Klassen einer High School, die überwiegend von Kindern der unteren und mittleren Mittelschicht besucht wird, so daß eine starke Streuungseinschränkung unwahrscheinlich ist. Daß in die Berechnung der Korrelation eine alterskorrigierte Variable, der IQ, und eine nicht alterskorrigierte Variable, das Kreativitätsmaß, eingingen, dürfte ebenfalls nicht zu einer starken Unterschätzung des Zusammenhangs geführt haben, da die Altersstreuung dieser Stichprobe geringer ist als z.B. diejenige des Samples von Yamamoto (1964), der für 272 Schüler 9., 1o., 11. und 12. Klassen einen Koeffizienten von .3o für die Kreativitäts-Intelligenz-Beziehung erhielt (vgl. Tabelle 1).
Insgesamt sprechen die in Tabelle 1 dargestellten Ergebnisse nicht dafür, daß die TTCT eine oder mehrere von der Intelligenz unabhängige Persönlichkeitsdimensionen messen. Torrance selbst hat die Enge des Zusammenhangs vor allem deshalb unterschätzt, weil er überwiegend Untersuchungen mit hochbegabten Schülern durchführte (vgl. S. 21). Eine mäßige oder sogar mittlere Korrelation der TTCT-Werte mit Intelligenztestergebnissen kann allerdings nicht auf jeden Fall als Hinweis auf die mangelnde Validität des Verfahrens gewertet werden; der Torrance-Test kann trotz dieses Zusammenhangs mit der Intelligenz der bessere Prädiktor realen kreativen Verhaltens sein. Auch daß die verschiedenen Kreativitätsmaße untereinander nicht höher als mit dem IQ korrelieren, müßte nicht unbedingt zu einem negativen Urteil über den Test führen (lediglich die Praxis, einen einzigen Kreativitäts-Gesamtwert zu berechnen, müßte aufgegeben werden), wenn die Untersuchungen zur Übereinstimmungs- und/oder Vorhersagevalidität zu befriedigenden Resultaten geführt hätten. Da in diesen Arbeiten jedoch vor allem Außenkriterien wie z.B. Schulleistungen, Lehrer- und Schülerurteile u.ä. verwendet wurden, die ebenfalls in erheblichem Ausmaß durch die Intelligenz determiniert sind, bleibt die Validitätsfrage trotz der großen Anzahl solcher Untersuchungen offen.

Versuche, sie zu beantworten, müssen sich daher auf die Resultate von Untersuchungen stützen, die von der Intelligenz unbeeinflußte Außenkriterien verwenden. Sie werden dadurch erschwert, daß Torrance keine theoretischen Aussagen über erwartete Zusammenhänge zwischen Kreativität und anderen Persönlichkeitsvariablen - abgesehen von der Intelligenz - macht, sondern unter der Überschrift "Konstruktvalidität" lediglich Untersuchungen referiert, die entweder Persönlichkeitscharakteristika von Individuen mit hohen und niedrigen TTCT-Ergebnissen vergleichen oder die Kreativitätstestscores mit anderen Persönlichkeitsmaßen korrelieren.

Die Untersuchungen mit Elementar- und jüngeren Sekundarschülern, die Torrance (1966, 1967a)als Belege für die Konstruktvalidität seines Tests anführt, sagen wesentlich mehr über verschiedene Persönlichkeitscharakteristika von Kindern mit hohen TTCT-Werten aus als über die Validität des Verfahrens. Es wird mitgeteilt, daß Einstellungsrigidität - gemessen mit einer revidierten Form des Fragebogens von Frenkel-Brunswik - hochsignifikant negativ mit dem Gesamtwert aus 4 Torrance-Aufgaben (r =-.41), mit Flüssigkeit (r =-.4o), Originalität (r =-.37) und Flexibilität (r =-.32) korreliert (Untersuchung von Fleming und Weintraub 1962 an 68 Elementarschülern), daß hochkreative Kinder ihre Meinung bei ungenügender Information nicht äußern, Ungewißheit leichter ertragen und voreilige Schlüsse vermeiden (Long und Henderson 1965, 288 Elementarschüler), daß Kinder mit hohen Originalitätswerten sich von Autoritätsfiguren und dem gleichgeschlechtlichen Elternteil distanzieren, weniger geachtet und abhängiger (!) sind und sich selbst als "unglücklich" bezeichnen (Long, Henderson und Ziller 1965, 111 Fünftkläßler), daß Spielfreude, Spontaneität,Humor und ähnliche Merkmale mit Flüssigkeit, Flexibilität und Originalität im Product Improvement Test signifikant korrelieren (Koeffizienten von .21 bis .36, Untersuchung von Lieberman 1965 an 93 Kindergarten-Kindern) und daß nach den TTCT hochkreative Kinder ein besseres Selbstbild, Humor, leichteren Zugang zu Erlebnissen der frühen Kindheit haben, nicht sehr stark verdrängte ödipale Ängste und eine ungleichmäßige Ego-Entwicklung i.S. von teilweise nicht altersgemäßem Verhalten zeigen (Weisberg und Springer 1961, 32 hochbegabte Viertkläßler). Torrance berichtet außerdem von einer eigenen Untersuchung (1962 a), in der festgestellt wurde, daß die nach dem TTCT-Ergeb-

nis kreativsten Kinder aus 23 Klassen hochoriginelle Zeichnungen und andere Produktionen verfertigten, daß sie bei Lehrern und Schülern für ihre ausgefallenen oder albernen Einfälle bekannt waren, und daß ihre kreativen Produkte humorvoll und spielerisch-entspannt wirkten. Die letzte der mit jüngeren Schülern durchgeführten Untersuchungen zur Konstruktvalidität, die Torrance in seinem Testmanual beschreibt, gibt Aufschluß über die sozialen Interaktionen von hochkreativen Kindern (hochkreativ nach den TTCT-Ergebnissen), die zusammen mit je 4 weniger kreativen Kindern an einer gemeinsamen Aufgabe arbeiteten. Die kreativen Kinder scheinen wenig Anerkennung für ihre zahlreichen Einfälle zu finden, sie werden im Gegenteil unter Druck gesetzt, die Vielzahl oder Originalität ihrer Beiträge zu reduzieren. Ihre Gegenmaßnahmen reichen von Nachgeben, Apathie, Ignorieren der Kritik, Clownerien über unerschütterliche Persistenz zu Gegenaggressionen und schwankenden Leistungen (Torrance 1963, zitiert nach Torrance 1966).
Sämtliche Ergebnisse werden ohne Darstellung und Diskussion des theoretischen Rahmens wiedergegeben, weniger günstige Befunde werden nicht mitgeteilt[1)], unerwartete Resultate - wie z.B. die Abhängigkeit der kreativen Kinder in der Untersuchung von Long, Henderson und Ziller (1965) - bleiben unerörtert, so daß es außerordentlich schwierig ist, aufgrund der Angaben von Torrance die Konstruktvalidität seines Tests zu beurteilen.
In einer Reihe weiterer Arbeiten wurde die Beziehung der TTCT-Ergebnisse zu verschiedenen Persönlichkeitsvariablen untersucht, deren Zusammenhang mit kreativem Verhalten bekannt oder nach theoretischen Annahmen wahrscheinlich war.
Hauck und Thomas (1972) fanden für 80 Viert-, Fünft- und Sechstkläßler eine sehr enge Beziehung zwischen einem TTCT-Gesamtwert (ohne nähere Angaben) und dem von den Mitschülern geschätzten Sinn für Humor; die Korrelation betrug .89 (diejenige zwischen Intelligenz und Humor .91 und die Kreativitäts-Intelligenz-Korrelation war .29). Die Untersuchungshypothese geht auf Ergebnisse

---

1) So fehlt beispielsweise bei der Beschreibung der Arbeit von Liebermann (1965) jeglicher Hinweis darauf, daß die Ratings für Spielfreude u.ä. (s.o.) mit einem Maß der allgemeinen Intelligenz noch etwas höher korrelierten als mit den TTCT-Ergebnissen und daß diese mit dem Intelligenzmaß etwa ebenso hoch korrelierten wie mit den Ratings.

von Getzels und Jackson (1962) und Torrance (1962a) zurück.

Eine bessere theoretische Fundierung einer Konstruktvalidierung erscheint beispielsweise mit Hilfe des Persönlichkeitsmerkmals Unabhängigkeit möglich, das als direkte Voraussetzung für das Finden neuer Lösungen betrachtet werden kann. Vor allem Crutchfield (1961, 1962, 1963) entwickelte eine Theorie, nach der sich Konformität und kreative Leistungen gegenseitig ausschließen. Er geht davon aus, daß in einer Problemsituation zwei unterschiedliche Motivationen im Individuum wirksam werden können, eine extrinsische, ichbezogene Motivation, die eine offene, unvoreingenommene Problemsicht und eine intensive sachbezogene Beschäftigung mit der Aufgabe unmöglich macht, oder eine intrinsische, rein aufgabenorientierte Motivation im Sinn eines starken Interesses an dem Problem selbst. In Experimenten wies Crutchfield nach, daß Konformitätsdruck extrinsische, ichnahe Motivationen erzeugt und daß es ein intraindividuell relativ konstantes Bedürfnis nach Konformität gibt, das nahezu inhalts- und problemunabhängig ist. Daß dieses Bedürfnis, sich konform zu verhalten, kreativen Leistungen abträglich ist, läßt sich aus diesen Überlegungen Crutchfields ableiten: ein konformes Individuum reagiert in einer Problemsituation mit einer starken Ichbeteiligung und versucht, seine Unsicherheit und Mißerfolgsangst dadurch zu reduzieren, daß es sich quasi des Schutzes seiner Gruppe versichert, indem es auf die in der Gruppe üblichen Verhaltensmuster zurückgreift. Ein solches Verhalten schließt jedoch kreative Problemlösungen aus.
Der aus der Theorie Crutchfields abzuleitende Zusammenhang zwischen Unabhängigkeit von der sozialen Umwelt und Kreativität war - z.T. schon vorher - von Barron (1953, 1961, 1969), Cattel und Drevdahl (1955), Gough (1961), Taylor und Barron (1963) u.a. an Erwachsenen festgestellt worden.
Wegen der guten theoretischen und empirischen Fundierung dieses Zusammenhangs wurden auch mehrere Untersuchungen zur Beziehung zwischen Unabhängigkeit und TTCT-Ergebnissen mit Kindern durchgeführt. Dacey und Ripple (1969) fanden jedoch ebenso wie Weisberg und Springer (1961) keine Beziehung zwischen Unabhängigkeit und den TTCT-Werten; in der von Torrance (1966) zitierten Untersuchung von Long, Henderson und Ziller wurde sogar ein positiver Zusammenhang zwischen Abhängigkeit und hohen Scores im Kreativitätstest festgestellt.

Bevor diese Ergebnisse als Hinweise auf die mangelnde Validität der TTCT gewertet werden können, wäre einmal die Altersentwicklung des Merkmals Unabhängigkeit zu überprüfen - eventuell zeigt es sich erst gegen Ende der Adoleszenz oder zu Beginn des Erwachsenenalters - und zum anderen müßten die in verschiedenen Untersuchungen verwendeten Operationalisierungen dieser Variablen daraufhin überprüft werden, ob sie wirklich das gleiche Konstrukt messen. (Den Angaben von Torrance zu der Untersuchung von Long, Henderson und Ziller sowie denjenigen von Dacey und Ripple und von Weisberg und Springer ist nicht zu entnehmen, wie Unabhängigkeit gemessen wurde.) Immerhin fanden Long und Henderson (1965), daß bei einer Stichprobe von 288 Kinder der 2. bis 7. Klasse die Unabhängigkeit der Meinungsbildung mit den Flüssigkeitswerten einer Torrance-Aufgabe signifikant positiv korrelierte und daß die Mittelwerte der Meinungsbildungsskala über die 6 Klassen hinweg nicht zunahmen. Leider blieb in dieser Untersuchung die mögliche Einflußvariable Intelligenz unberücksichtigt. Werner (1966) und Werner und Bachtold (1969) charakterisierten die hochbegabten Kinder und Jugendlichen ihrer Untersuchungen ebenfalls als unabhängig (nach dem Ergebnis des Children's Personality Questionnaire von Porter und Cattel).
Somit kann nur gesagt werden, daß Untersuchungen mit der Persönlichkeitsvariable Unabhängigkeit bisher nicht zu Nachweisen der Konstruktvalidität der TTCT führten.

Der Zusammenhang der Ergebnisse der TTCT mit einer Variable des kognitiven Stils, der Feldabhängigkeit-Feldunabhängigkeit, wurde im Rahmen der Konstruktvalidierung ebenfalls untersucht. Ursprünglich als Charakteristikum des Wahrnehmungsverhaltens betrachtet, zeigte sich nach Witkin (1973), daß Feldabhängige auch bestimmte Formen des Denkens und Problemlösens aufweisen, die denjenigen feldunabhängiger Individuen entgegengesetzt sind. Darüber hinaus korreliert die Feldunabhängigkeit auch mit Charakteristika des Sozialverhaltens. Witkin betont, daß sein Konzept des "kognitiven Stils" perzeptuelle, kognitive und Persönlichkeitsvariablen umfaßt.[1]. Seine Theorie über die Entstehung des kognitiven Stils der Feldabhängigkeit bzw. -unabhängigkeit stellt zugleich die Beziehung dieser Variablen zu kreativem Verhalten her.

---

1) Witkins Konstrukt der Feldabhängigkeit wird allerdings z.T. stark kritisiert; Schulte (1974) lehnt es z.B. aus theoretischen und methodischen Gründen und aufgrund seiner eigenen Untersuchungsergebnisse ab.

Witkin vertritt die Ansicht, daß Umwelteinflüsse im Vergleich zu genetischen Faktoren auf jeden Fall für den weitaus größeren Varianzanteil dieses Merkmals verantwortlich zu machen sind. Unter ersteren spielen die Mutter-Kind-Interaktionen die wichtigste Rolle; ein Kind wird umso weniger feldabhängig, je mehr es zu autonomem, unabhängigem Verhalten ermutigt wurde. Witkins Auffassung wurde durch eine Reihe von Untersuchungen in den USA und mehreren anderen Ländern bestätigt, so z.B. durch eine Untersuchung in je 2 holländischen, italienischen und mexikanischen Dörfern, die sich jeweils im Ausmaß der Bedeutung unterscheiden, die sie der Konformität gegenüber Familie und religiösen und politischen Autoritäten beimessen.[1] Der nachgewiesene enge Zusammenhang zwischen Erziehung zu konformem Verhalten und Feldabhängigkeit führte mehrere Untersucher dazu, eine ebenfalls enge Beziehung zwischen Feldunabhängigkeit und Kreativität zu vermuten.[2] Die Ergebnisse einiger, aber nicht aller Untersuchungen mit Erwachsenen bestätigen diese Vermutung.[3]

Untersuchungen über den Zusammenhang zwischen Feldabhängigkeit und der mit den TTCT gemessenen Kreativität, die mit Kindern durchgeführt wurden, erbrachten ebenfalls keine eindeutigen, aber überwiegend negative Resultate.[4]

Obwohl schon relativ früh eine negative Korrelation zwischen allgemeiner Intelligenz und Feldabhängigkeit festgestellt wurde[5], und obwohl inzwischen nachgewiesen wurde, daß auch letztere eine stark altersabhängige Persönlichkeitsvariable ist[6], scheinen bislang Untersuchungen zu fehlen, die die Einflußgrößen Alter und Intelligenz angemessen kontrollieren.

Somit dürfte die Persönlichkeitsdimension Feldabhängigkeit-Feldunabhängigkeit auf jüngeren Altersstufen noch zu wenig untersucht sein, um sie als Kriteriumsvariable bei der Validierung eines Kreativitätstests einsetzen zu können. Aufgrund der erwähnten Unter-

---

1) Witkin, Price-Williams et al. 1974

2) Rookey und Reardon (1973) gehen sogar soweit, als einziges Verfahren zur Messung kreativer Fähigkeiten den "Hidden Figures Test" von Fanani (1964) zu verwenden.

3) Positive Korrelationen zwischen beiden Variablen fanden Bieri et al. 1958, Mac Kinnon 1960 (zitiert nach Christie 1970); Spotts and Mackler 1967; Bloomberg (1971) fand seine Hypothese, daß in einer Gruppe kreativer Studenten die feldunabhängigen überwiegen, nicht bestätigt.

4) Z.B. Mc Whinnie (1967, 1969, 1970), Stevens (1970)

5) Goodenough and Karp 1961, Hardesty and Priester 1956, beide zitiert nach Seiffge-Krenke 1974

6) Bigelow 1971, Wolf 1971

suchungen ist daher auch kein Urteil über die Konstruktvalidität der TTCT möglich.

Eine abschließende Beurteilung der Konstruktvalidität des Kreativitätstests von Torrance kann daher trotz der großen Zahl von Validitätsuntersuchungen lediglich feststellen, daß die Mehrzahl der bisherigen Ergebnisse kein günstiges Urteil erlaubt.
Crockenberg (1972) formuliert etwas vorsichtiger: "Clearly, while the validity evidence on the Torrance is abundant it is also weak. While we cannot readily dismiss the Torrance Test, neither is it reasonable to conclude that it is a valid maesure of creative process". (S. 35)

2.23 Ausgewählte Untersuchungsergebnisse

Im Verlauf der Entwicklung oder Adaptierung eines Tests zur Erfassung kognitiver Voraussetzungen für kreatives Verhalten wird eine Reihe von Fragen auftreten, für deren Beantwortung die vorliegenden Untersuchungsergebnisse der Kreativitätsforschung mit Kindern nutzbar gemacht werden sollen. Vor allem folgende Fragen sollen aufgrund möglichst umfassender Informationen beantwortet werden:

- Fragen zur Testsituation
- Fragen zur Testauswertung
- Bewertung eventueller Abhängigkeiten der Testergebnisse vom Geschlecht und/oder dem sozio-ökonomischen Status der Probanden
- Fragen zum Zusammenhang zwischen Kreativität, Intelligenz und Schulleistung.

Deshalb werden die mit dem jeweiligen Meßverfahren durchgeführten Untersuchungen zu diesen Themen dargestellt und diskutiert.

2.231 Auswirkungen unterschiedlicher Testbedingungen

Der Kreativitätstest von Torrance ist mehrfach wegen seiner testähnlichen Untersuchungssituation kritisiert worden[1]; Einwände richteten sich vor allem gegen die Zeitbegrenzung und den durch die Instruktion hervorgerufenen Leistungsdruck (vgl. S.17 - 18).

---

1) Crockenberg 1972, Krause 1972, Vernon 1971, Wallach and Kogan 1965, Wallach 197o

Eine Reihe von Untersuchungen überprüfte die Auswirkungen von Änderungen der Zeitbegrenzungen und der Testsituation.
Während auf dem Gebiet der Intelligenzmessung bereits relativ früh systematische Untersuchungen über differentielle Auswirkungen unterschiedlicher Zeitbegrenzungen mit Intelligenztests durchgeführt wurden[1], liegen für Kreativitätstests nur einige kleinere Arbeiten vor, die diese Frage auf sehr verschiedene Arten angehen, so daß generalisierende Aussagen noch nicht möglich sind. Torrance (1972a) stellt eine Anzahl von Untersuchungen über Auswirkungen verschiedener Testbedingungen zusammen, in der auch 4 Arbeiten enthalten sind, die Effekte unterschiedlicher Testzeiten auf die Ergebnisse der TTCT prüften. Während nach Khatena (1971 b, zitiert nach Torrance 1972a)Variationen der Zeitbegrenzungen bei Schülern 1o. und 12. Klassen zu signifikanten Unterschieden in der Testleistung führten - eine verlängerte "Inkubationszeit" führte zu höheren Originalitätswerten[2] -, erzielte Towell (1972, ebenfalls zitiert nach Torrance 1972a)bei Wegfall der Zeitbegrenzung keine signifikante Verbesserung der Testleistungen von Viertkläßlern. In den beiden anderen Untersuchungen ist die Testzeitvariable mit anderen Variablen der Testsituation konfundiert, so daß nicht klar ist, ob erstere der ausschlaggebende Faktor war; beide Arbeiten deuten darauf hin, daß eine Bearbeitung der Tests zu Hause, also ohne Aufsicht und Zeitbegrenzung, validere Ergebnisse erbringt als das Standardverfahren (Torrance 1969 mit 3o hochbegabten Sechstkläßlern) bzw. als das Standardverfahren sowie eine Testsituation, die den Schülern eine Inkubationszeit zugestand, und eine spielähnliche Testsituation (Van Mondfrans et al. 1971 mit 319 Schülern 5., 8. und 11. Klassen). Van Mondfrans et al. führen die günstigen Ergebnisse der Testbearbeitung zu Hause auf den Wegfall der Zeitbegrenzung zurück.

Die Arbeiten zur Auswirkung der Testatmosphäre verwendeten als unabhängige Variablen spielerische, entspannte, "sichere" oder ungezwungene versus testähnliche, stressinduzierende oder wettbewerbsähnliche Untersuchungssituationen. Auch hier waren die Variablen der Testsituation z.T. mit denjenigen der Testzeit konfundiert.

---

1) S. dazu Bennett 1941, Davidson and Carroll 1945, Freeman 1966

2) Ob sich dadurch die Validität erhöht, ist ungewiß. Eine gleichmäßige Erhöhung der Testergebnisse aller Pbn würde jedenfalls keine Veränderung der Testgütekriterien bewirken.

Boersma und O'Bryan (1968) führten eine Untersuchung an 46 Jungen der 4. Klassenstufe durch, in der die Auswirkungen der Standard-Testsituation (in Anwesenheit des Lehrers) und einer völlig entspannten und für die Schüler anscheinend freiwilligen Testung außerhalb der Schule auf die Ergebnisse zweier Torrance-Aufgaben - Figure Completion und Unusual Uses - überprüft wurden. Beide Gruppen bearbeiteten die Aufgaben in der vorgeschriebenen Zeit (bei der "Freiwilligen"-Gruppe fragte der Testleiter nach Ablauf der Testzeit der ersten Aufgabe, wer mitkommen wolle, um das Schwimmbad zu besichtigen!). Die Kinder, die in der ungezwungenen, entspannten und schulunähnlichen Situationen getestet wurden, erreichten in der verbalen und in der figuralen Torrance-Aufgabe signifikant höhere Ergebnisse als die gleich intelligente Gruppe, die unter Standardbedingungen getestet wurde. Die Beziehung zwischen Intelligenz- und Kreativitätsmaßen war in der "freiwilligen" Gruppe insignifikant, z.T. sogar negativ, in der "Standard"-Gruppe waren 3 von 4 Koeffizienten signifikant oder hochsignifikant.

Vernon (1971) testete etwa 4oo Achtkläßler mit verschiedenen Kreativitätstestaufgaben von Torrance und von Wallach und Kogan in 2 verschiedenen Testsituationen, der Standard-Prozedur - allerdings nach 2 oder mehr Testungen mit anderen Verfahren! - und einer entspannten, spielähnlichen Situation ohne Zeitbegrenzung für die einzelnen Aufgaben. Insgesamt erreichten die Schüler der zweiten Gruppe höhere Flüssigkeits- und Originalitätswerte, und ihre Scores zeigten bei den meisten Aufgaben eine stärkere Streuung. Die durchschnittlichen Interkorrelationen der 7 Tests waren .346 in der "entspannten" und .287 in der "Standard"-Gruppe, aber die Korrelationen der Kreativitätstestscores mit den Ergebnissen verschiedener Intelligenztests waren bei ersteren sogar etwas höher - allerdings nicht signifikant - als bei letzteren. Somit waren sämtliche Korrelationen der unter entspannten Bedingungen erzielten Testergebnisse höher, was eventuell auf die größere Streuung dieser Kreativitätsscores (vgl.o.) zurückzuführen ist; die Hypothese von Wallach und Kogan (1965), daß unter entspannten, spielähnlichen Testbedingungen ohne Zeitbegrenzung die Kreativitäts-Intelligenz-Korrelation niedriger ist als in der normalen Testsituation, wurde in dieser Untersuchung jedenfalls nicht bestätigt. Der Grund dafür ist eventuell in der etwas zu unstrukturierten Testsituation Vernons zu suchen.

Vernon berichtet, daß die permissive Atmosphäre dazu führte, daß die Schüler miteinander sprachen und z.T. voneinander abschrieben und daß einige Schüler zu lange an bestimmten Aufgaben arbeiteten und dann nicht mehr genügend Zeit für die übrigen Items hatten (für die gesamte Testung standen 2 Schulstunden zur Verfügung). Wallach und Kogan führten dagegen Individualtestungen ohne jegliche Zeitbegrenzung durch, so daß den Schülern für alle Aufgaben genügend Zeit blieb. Darüber hinaus wurde durch die direkte Konfrontation mit dem Testleiter wahrscheinlich trotz der entspannten, spielerischen Atmosphäre ein gewisser Druck auf die Kinder ausgeübt, möglichst viele Antworten auf jedes Item zu geben, so daß die Testsituation Wallachs und Kogans besser strukturiert und kontrolliert war als diejenige von Vernon.
Wichtiger als die nicht im gewünschten Ausmaß gelungene diskriminante Validierung erscheint das Ergebnis, daß die unter entspannten Testbedingungen erzielten Kreativitätstestscores engere Zusammenhänge mit einer Reihe von Außenkriterien - wie z.B. künstlerische Freizeitinteressen, soziometrische Nominierungen zu der Frage nach einer kreativen Betätigung, Lehrerratings der Neugier der Pbn u.ä. - aufwiesen als die unter Standardbedingungen erreichten Scores.
Interessant erscheinen auch in diesem Zusammenhang die Ergebnisse der Untersuchung von Van Mondfrans et al. (1971), die u.a. eine spielähnliche Testsituation mit Zeitbegrenzung in ihr Experiment einbezogen (vgl. S. 56). Mit dieser Testbedingung wurden die schlechtesten Ergebnisse im Hinblick auf die Interkorrelationen der Kreativitätstestwerte und ebenso hohe Korrelationen dieser Scores mit den Intelligenztestergebnissen wie unter Standardbedingungen erzielt. Ob die Kinder die Testsituation wirklich als spielähnlich wahrnahmen, erscheint allerdings zweifelhaft, da der einzige Unterschied zum Standardverfahren in einer Abänderung der Instruktion bestand. Eventuell war in dieser Untersuchung die Situation zu schulähnlich, um in den Kindern eine entspannte und zugleich aufgabenorientierte Arbeitshaltung zu erzeugen.
Seiffge-Krenke (1974) führte ein Experiment mit 4 verschiedenen Testbedingungen durch, in dem eine Konfundierung verschiedener Variablen weitgehend vermieden wurde. Sie teilte 88 Viertkläßler auf folgende Gruppen auf: Stress mit und ohne Zeitbegrenzung, spiele-

rische Testsituation mit und ohne Zeitbegrenzung. Die Kinder in den 4 Gruppen waren nach Geschlecht, Alter und Intelligenz parallelisiert. Die besten Ergebnisse erzielten die Pbn, die in der spielähnlichen Testsituation ohne Zeitbegrenzung, die schlechtesten diejenigen, die unter Stress und mit Zeitbegrenzung getestet wurden. Die Auswirkungen der verschiedenen Testsituationen auf die Validität der TTCT-Ergebnisse wurde nicht untersucht, es ist jedoch anzunehmen, daß auch bei Kreativitätstests eine durch verbesserte motivationale Bedingungen gesteigerte Test-Performanz validere Rückschlüsse auf die dahinterstehende Kompetenz ermöglicht. Die Verbesserung der Motivation wurde von Seiffge-Krenke vermutlich durch Schaffung einer schulunähnlichen Situation erreicht: während in den Stress-Situationen relativ autoritär Leistungen gefordert, eine Wettkampfatmosphäre geschaffen und das Interesse des Lehrers an den Ergebnissen hervorgehoben wurden, betonte die Instruktion der spielerischen Bedingung den Spaß, den die Aufgaben machen.
Zwei Untersuchungen mit Studenten, die die Effekte verschiedener Testsituationen auf TTCT-Ergebnisse prüften, sind ebenfalls weniger aufschlußreich, da auch sie keine Aussagen über die Validität der unter beiden Bedingungen erzielten Resultate gestatten. Dentler und Mackler (1964) fanden, daß eine durch die freundliche, ermutigende Art des Testleiters geschaffene "sichere" Testatmosphäre im Vergleich mit 3 unterschiedlichen, aber sich gleich auswirkenden Kontrollbedingungen zu sehr viel höheren Originalitätsscores im Unusual Uses Test führte.

Die Untersuchungshypothese von Bader (197o) wurde nicht bestätigt: von den 3 Stufen seiner unabhängigen Variable - einer entspannten, ungezwungenen, einer neutralen und einer formalen, wettbewerbsbetonten Testsituation - förderte nicht die erstgenannte, sondern die neutrale Bedingung die Ideenproduktion in der verbalen Form der TTCT am meisten; die beiden anderen Testsituationen hatten in etwa den gleichen Effekt. (Versuchspersonen waren 9o Studenten.) Insgesamt betrachtet scheinen die Torrance-Tests bei jüngeren Schülern unter entspannteren, weniger schulähnlichen Testbedingungen validere Ergebnisse zu erbringen als unter Standardbedingungen. Zugleich erscheint es jedoch erforderlich, die Testsituation so zu strukturieren, daß bei den Kindern eine aufgabenorientierte Einstellung erzeugt und aufrecht erhalten wird.

Über die Auswirkungen verschiedener Testzeiten oder den Wegfall der Zeitbegrenzung erscheinen aufgrund der hier dargestellten Untersuchungsergebnisse keine Aussagen möglich. Dieses Problem und die Frage nach der günstigsten Testatmosphäre werden bei der Besprechung der Untersuchungsergebnisse zu dem Kreativitätstest von Wallach und Kogan noch einmal aufgegriffen.

2.232 Testauswertung

Die von Torrance entwickelte Auswertung seiner Aufgaben nach Flüssigkeit, Flexibilität, Originalität und Elaboration war wegen der zu erwartenden Zusammenhänge zwischen diesen Maßen kritisiert worden (vgl. S. 19 -2o). Die meisten dieser Erwartungen wurden durch Untersuchungsergebnisse zu den Interkorrelationen der Maße bestätigt (s.S.3o -31). Auch die Ergebnisse einiger Faktorenanalysen erbrachten keine Begründung für die Auswertung nach 4 verschiedenen Kriterien: Cicirelli (1965) reduzierte aufgrund der Faktorenstruktur der 4 verbalen und der 4 figuralen Maße der TTCT die Auswertung auf 4 Maße, verbale Flüssigkeit-Flexibilität-Originalität, verbale Elaboration, figurale Flüssigkeit-Flexibilität-Originalität und figurale Elaboration (Untersuchung mit 6o9 Sechstkläßlern). Madaus (1967) führte eine Faktorenanalyse mit denselben 8 Maßen und dem verbalen und numerischen Gesamtwert des School and College Ability Test (SCAT) durch und erhielt nach einer Varimax-Rotation einen "Intelligenz-Faktor", auf dem der verbale SCAT-Gesamtwert und der Score für verbale Elaboration luden, einen verbalen Kreativitätsfaktor, auf dem außer den übrigen 3 verbalen Maßen auch derjenige für figurale Elaboration lud, und einen figuralen Kreativitätsfaktor, auf dem neben den figuralen Flüssigkeits-, Flexibilitäts- und Originalitätsscores auch der numerische SCAT-Gesamtwert eine sehr hohe Ladung hatte (Probanden waren 6o9 Neunt- und Zehntkläßler). Nach beiden Untersuchungen messen die beiden Elaborationsscores etwas anderes als die 3 übrigen verbalen und die 3 figuralen Maße. Harvey, Hoffmeister, Coates und White (197o) werteten die 5 verbalen Torrance-Aufgaben nach Flüssigkeit, Flexibilität und Originalität aus und faktoranalysierten die resultierenden 2o Testwerte[1)] von 64 erwachsenen Pbn; sie erhielten 2 aufgabenspezifische Faktoren (Ask-and-Guess und Unusual Uses), einen Flüssigkeits-Flexibilitäts- und einen Originalitäts-Faktor.

---

1) Für die 3teilige Aufgabe "Ask-and-Guess" ergaben sich 9 Scores, die Aufgabe "Unusual Questions" wurde nicht nach Flexibilität ausgewertet.

Nach diesen Ergebnissen dürfte feststehen, daß zumindest ein Kreativitätsmaß, Flüssigkeit oder Flexibilität, redundant ist, bei Kindern vermutlich sogar zwei, da diese beiden Maße in 2 Faktorenanalysen zusammen mit Originalität auf einem Faktor luden. Ein Urteil darüber, welche Maße zur Messung kreativer Fähigkeiten beitragen und welche nicht, soll aufgrund der Ergebnisse der Validitäts- und Retest-Reliabilitätsuntersuchungen gefällt werden.
Die meisten der von Torrance (1966, 1967a) mitgeteilten Ergebnisse werden nicht für die verschiedenen Kreativitätsmaße getrennt aufgeführt. In den Untersuchungen zur Übereinstimmungsvalidität schneiden nach Torrance (1966) bei der Verwendung des Außenkriteriums Lehrerurteil Flüssigkeit[1] oder Flüssigkeit und Originalität[2] am besten ab (da weitere Angaben fehlen, wird angenommen, daß es sich um Gesamtwerte aus verbalen und figuralen Maßen handelt); Nelson (1963, zitiert nach Torrance 1966) fand für die 4 verbalen Maße etwa gleich gute Ergebnisse, während von den figuralen Maßen nur Originalität und Elaboration zwischen den aufgrund des Lehrerurteils gebildeten Extremgruppen differenzierten.
In der Untersuchung von Seiffge-Krenke wurde zwar mit Hilfe der Originalitätswerte die geringste Zahl von Fehlklassifikationen in die aufgrund des Lehrerurteils und in die aufgrund der Schülerratings gebildeten Extremgruppen erzielt, aber in beiden Fällen waren die Flüssigkeits- und Elaborationsscores die Maße, die in der Diskriminanzanalyse die höchsten F-Werte erreichten.
Bei der Darstellung der Untersuchungen, die Schulleistungen als Validitätskriterium verwenden, macht Torrance (1966, 1967a) keine nach den verschiedenen Maßen differenzierten Angaben, sondern bemerkt lediglich, daß die Elaborationswerte relativ enge Beziehungen zu den Schulleistungen hatten. Da jedoch Schulleistungen ein sehr zweifelhaftes Validitätskriterium für einen Kreativitätstest sind, soll dieser Hinweis unberücksichtigt bleiben.

In der Untersuchung zur prädiktiven Validität der TTCT von Erickson (1966, zitiert nach Torrance 1966) erzielte der Gesamtwert für Flüssigkeit als einziger eine auf dem 5 Prozent-Niveau signifikante Korrelation mit dem Kriterium (vgl. S. 42-43). Nach Eliminierung verschiedener irrelevanter Fragebogenitems und einer Umwandlung dieses Kriteriums in 3 Kriteriumsmaße - Qualität und Quantität kreativer Leistungen und Zukunftspläne - erhielt Torrance (1969 a und b, zitiert nach Torrance 1972b) nach 7 Jahren für 46 der 69 ehemaligen High

---

1) Yamamoto 1962, zitiert nach Torrance 1966

2) Torrance and Gupta 1964, zitiert nach Torrance 1966

School Senioren überwiegend hochsignifikante Korrelationen mit den 4 Kreativitätsmaßen (.25 ≷ r ≷ .48, Md=.41), wobei der Flexibilitätsscore am besten, der Elaborationswert am schlechtesten abschnitt.

Dieselben Kriterien wurden 1971 an den Schülern erhoben, die 1959 die 7. bis 12. Klassen der High School der Minnesota University besucht hatten; 236 der 392 ehemaligen Schüler beantworteten die Fragebogen. Für diese Stichprobe waren die Originalitätswerte die besten und die Flexibilitätswerte die schlechtesten Prädiktoren. Eine Aufschlüsselung der Originalitätskorrelationen mit den Kriterien nach den 6 Klassenstufen zeigt allerdings, daß die Vorhersagen für die 2 untersten Klassen, also für die 7. und 8. Klasse, sehr schlecht waren, von 6 Koeffizienten war nur einer signifikant (.14 ≷ r ≷ .41, Md =.235).
Es ist anzunehmen,daß auf jüngeren Altersstufen alle übrigen Auswertungsmodi schlechter abschneiden als Flüssigkeit, da letztere nur ein Auszählen der akzeptablen Antworten erfordert, während die anderen 3 Auswertungskriterien trotz der detaillierten Auswertungsrichtlinien letztlich Ratings von den Auswertern verlangen, jedenfalls für die Antworten, die nicht in den Richtlinien enthalten sind. Solche Ratings fallen den Auswertern jedoch wahrscheinlich umso schwerer, je größer der Altersabstand zu den Probanden ist. Somit können die hier dargestellten Ergebnisse für höhere Klassenstufen vermutlich nicht auf jüngere Schüler verallgemeinert werden.
Sofern in den Angaben zur Konstruktvalidität Korrelationskoeffizienten für die verschiedenen Kreativitätsmaße mitgeteilt werden, ergeben sich die höchsten Werte häufig für das Flüssigkeitsmaß [1]; eine Ausnahme bilden die Resultate der Untersuchungen zur Beziehung zwischen Feldabhängigkeit und Kreativität mit ihren insgesamt inkonsistenten Ergebnissen.[2]

Aus den Ergebnissen dieser Validitätsuntersuchungen kann nicht geschlossen werden, daß eines der mit größerem Auswertungsaufwand verbundenen Kreativitätsmaße validere Meßwerte liefert als der auf der Anzahl der akzeptablen Lösungen beruhende Flüssigkeitsscore.

---

1) In der Untersuchung von Fleming und Weintraub 1962 und Long und Henderson 1965

2) Vgl. S. 53 - 54

Auch bei den Untersuchungen zur Retest-Reliabilität schneiden die Flüssigkeitsmaße, vor allem diejenigen für verbale Flüssigkeit, am besten ab.[1)]

Demgegenüber gewinnt man aus Torrances Beschreibungen seiner Auswertungskriterien den Eindruck, als betrachte er Flüssigkeit in der Produktion von Einfällen als ein weniger valides Maß kreativer Fähigkeiten als Originalität[2)] und Flexibilität. So schreibt er z.B. über den figuralen Flüssigkeitsscore: "This score is useful primarily in helping the user understand the other figural scores. The impulsive thinker, the banal thinker, and even the "non-thinker" can achieve rather easily high scores, even maximum scores within the time limits." (Torrance 1966, S. 74). Das Originalitätsmaß wird dagegen für wesentlich eindeutiger gehalten: "The measure of originality comes closest to the essence of creativity." (Torrance, Tan und Allman 197o, S. 336). Daß die Auswertungsobjektivität für beide Originalitätsscores niedriger ist als für alle anderen verbalen und figuralen Maße[3)] und daß auch die detaillierteste Auflistung statistisch häufiger und damit unoriginellerer Antworten nicht für alle Altersstufen vom Kindergarten bis zum College-Alter und in verschiedenen Ländern gleich valide sein kann, wird nicht erörtert[4)]. Außerdem müßte eine solche Liste häufig auftretender Antworten in regelmäßigen Zeitabständen von höchstens einem halben Jahr überprüft und ggf. abgeändert werden, da im Zeitalter der Massenmedien bestimmte Vorstellungen und Begriffe ganz plötzlich wesentlich häufiger auftreten können (man denke nur an die Auswirkungen beliebter Kinderserien im Fernsehen!) und damit "originelle" Antworten zu üblichen werden.

Von den 4 Kreativitätsmaßen zeigten die Elaborationsscores als einzige keine hohen Interkorrelationen mit den übrigen Maßen.[5)] Es wäre somit möglich, daß es sich bei diesen Werten um nicht redundante Kreativitätsmaße handelt. Einige Untersuchungsergebnisse

---

1) S. Torrance 1966, 1967a, Dewing 197o a, Wodtke 1964

2) In mindestens 2 Untersuchungen wurde der Originalitätsscore als einziges Maß kreativer Fähigkeiten verwendet, vgl. Torrance 1963b, und Torrance, Tan und Allman 197o.

3) Vgl. Torrance 1966, S. 19

4) Über die Stichproben, aus deren Antworten die Listen erstellt wurden, und die dabei angewandte Methode wird kaum etwas mitgeteilt. Torrance (1962b)macht einige Angaben, aber zu diesem Zeitpunkt lagen die verschiedenen Auswertungskriterien noch nicht in der heutigen Form vor.

5) Vgl. S. 6o - 61

lassen jedoch die Vermutung aufkommen, daß die Elaborationsscores eher schulleistungsrelevante Fähigkeiten erfassen als intellektuelle Voraussetzungen für kreative Leistungen. So lud in der Untersuchung von Madaus (1967) der verbale Elaborationswert zusammen mit dem verbalen Gesamtwert des School and College Ability Test auf einem Faktor; die Ladungen waren beide sehr hoch - um .90 - und kein anderes der 8 Kreativitätsmaße zeigte eine Ladung über .09 auf diesem Faktor. In der Untersuchung von Cicirelli (1965) sagte der verbale Elaborationsscore die Schulleistungen in Sprache und Lesen besser vorher als das auf Flüssigkeit, Flexibilität und Originalität basierende verbale Gesamtmaß, und der figurale Elaborationswert sagte die beiden genannten Schulleistungen und die Mathematikleistungen besser vorher als der figurale Flüssigkeits-Flexibilitäts-Originalitätswert. Nach Auspartialisierung der Intelligenz waren die beiden Elaborationsmaße immer noch die besseren Prädiktoren der Schulleistungen in Sprache und Lesen. Auch Torrance (1966) weist auf die positive Beziehung der Elaborationswerte zu Schulleistungen hin.
Schließlich werden auch durch die Untersuchungsergebnisse von Dauw (1965, zitiert nach Torrance 1966, 1967a)High-School-Senioren mit hohen Elaborationswerten als Individuen charakterisiert, die weniger kreatives Verhalten als gute Schulleistungen erwarten lassen. Aufgrund einer Reihe von Fragebogenergebnissen hatten diese Schüler im Vergleich mit Schülern, die hohe Originalitätswerte erzielt hatten, weniger kreative und konventionellere Berufswünsche, die zugleich weniger intensiv waren, bessere Noten, Befürchtungen, die von Eltern und Lehrern in sie gesetzten Erwartungen nicht zu erfüllen, und sie gingen in Problemsituationen und bei Stress nach ihren Angaben häufiger aus dem Feld. An Sozialkontakten zeigten sie sich interessierter als die Schüler mit hohen Originalitätswerten.

Aufgrund dieser Untersuchungsergebnisse erscheint es nicht angezeigt, von einer Auswertung der TTCT nach dem Kriterium Elaboration valide Informationen über kreative Fähigkeiten zu erwarten.

Eine abschließende Beurteilung aller Untersuchungsergebnisse zu der Frage der verschiedenen Auswertungskriterien der Torrance-Tests bestätigt das Fazit, das Hargreaves und Bolton (1972) aus ihren Resultaten ziehen: "...the time and effort required to calculate anything more than Fluency scores on divergent tests do not justify the small amount of extra information gained." (S. 460)[1)]

---

1) Verständlicher wird das Zitat, wenn "do not justify" durch "are not justified by" ersetzt wird.

2.233 Geschlechtsabhängigkeit der TTCT-Ergebnisse

Torrance (1965a) vertritt die Ansicht, daß geschlechtsspezifische Leistungsunterschiede in vielen Verhaltensbereichen eine Folge der unterschiedlichen Belohnungen sind, mit denen die soziale Umwelt auf die jeweiligen Leistungsbemühungen von Jungen und Mädchen reagiert. Im Bereich der Schule zeigen die Mädchen nach seiner Meinung mehr erwünschte Verhaltensweisen als die Jungen und erhalten daher im normalen Unterricht auch mehr Belohnungen.[1] Wenn Lehrer aufgefordert worden waren, kreatives Verhalten - in Form ungewöhnlicher Einfälle und Fragen u.ä.- zu belohnen und anschließend über die Schülerbeiträge befragt wurden, zeigte sich, daß sie signifikant häufiger berichteten, Jungen belohnt zu haben. Auch die Schüler selbst nominierten in zwei Untersuchungen häufiger Jungen als Mädchen auf Fragen nach den besseren Einfällen im Umgang mit wissenschaftlichem Spielmaterial und nach der Überlegenheit beim Erklären wissenschaftlicher Prinzipien, obwohl sich in einer der beiden Untersuchungen die Leistungen von Jungen und Mädchen nicht unterschieden. Torrance hat durch diese kleineren Untersuchungen die im Bereich der Familienerziehung intensiv erforschte und gut belegte Wirksamkeit der geschlechtsrollenspezifischen Erwartungen und Belohnungen auch auf dem Gebiet der schulischen Erziehung demonstriert.

Die Frage ist nun, ob die Geschlechtsspezifität kreativer Verhaltensweisen bei Schulkindern bereits so stark internalisiert ist, daß sie auch auf neue kreativitätsrelevante Situationen, für die noch keine Geschlechtsrollen-Standards vorliegen, generalisiert. Mit anderen Worten: leisten Mädchen in Kreativitätstests oder jedenfalls in bestimmten Aufgaben weniger als Jungen, weil sie für reales kreatives Verhalten weniger Belohnungen erhielten? Zeigt sich der Unterschied schon zu Beginn der Schulzeit oder erst später, was eventuell bedeutet, daß neben dem motivationalen Unterschied eine Ungleichheit in den Lern- und Übungsgelegenheiten wirksam geworden ist?

Eine ganze Reihe von Untersuchungen hat u.a. die Geschlechtsabhängigkeit der TTCT-Ergebnisse miterfaßt, die relevanten Resultate werden in Tabelle 2 zusammengefaßt.

---

1) Belohnungen in Form von Schulerfolg werden den Mädchen auch in der BRD häufiger zuteil: Sie haben bessere Noten und zwar sowohl in Elementar- und Hauptschulen (s. Starck 1974) als auch in Mittelschulen (s.z.B. Amelang und Vagt 1970) und Gymnasien (s. Knoche 1969); in allen Schultypen wiederholen Jungen häufiger eine Klasse als Mädchen (s. Starck 1974).

Tabelle 2: Untersuchungsergebnisse zur Geschlechtsabhängigkeit der TTCT-Werte

Autor(en)	Stichprobe	Ergebnis
Fleming und Weintraub 1962	68 hochbegabte Elementarschüler	Keine signifikanten Unterschiede zwischen Jungen und Mädchen in dem verbalen und figuralen Gesamtwert
Yamamoto 1962, zitiert nach Torrance 1965a	Schüler 1. - 12. Klassen und Studenten (keine N-Angaben)	Bessere Ergebnisse der weiblichen Pbn im Ask-and-Guess-Test, außer in der 2., 3. und 4. Klasse, in denen die Jungen signifikant besser waren
Torrance 1963, zitiert nach Arasteh 1968	Kinder bis zu 5 Jahren (ohne nähere Angaben)	Keine Unterschiede in den TTCT-Ergebnissen zwischen Jungen und Mädchen
Yamamoto 1964b	272 Schüler 9. - 12. Klassen	Kein Geschlechtsunterschied im Gesamtwert aus 6 Kreativitätsmaßen für 2 Torrance-Aufgaben
Yamamoto 1965b	461 und 827 Fünftkläßler	Keine Geschlechtsunterschiede im Gesamtwert aus 5 Kreativitätsmaßen für 2 Torrance-Aufgaben
Torrance 1965a	5o hochbegabte Sechstkläßler	Jungen waren signifikant besser in der figuralen Originalität, Mädchen siginifikant besser in figuraler Elaboration und in einer verbalen Aufgabe
Torrance 1965a	75 hochbegabte Schüler 7. und 8. Klassen	Jungen waren in allen figuralen Maßen und dem verbalen Flexibilitätsmaß eines Tests besser, alle übrigen Unterschiede waren nicht signifikant
Long und Henderson 1965	288 Schüler 2.- 7.Klassen	Mädchen waren in der figuralen Flüssigkeit des "Lines"-Test (Parallelaufgabe zu "Circles") besser als Jungen
Cicirelli 1967	8o Sechstkläßler aus Familien mit 2 Kindern	Kein Unterschied in den verbalen und figuralen Flüssigkeits-Flexibilitäts-Originalitäts- und Elaborations-Werten
McWhinnie 1968	2 Gruppen (64 und 48) Sechstkläßler	Jungen der Gruppe 1 waren signifikant besser als die Mädchen in figuraler Flüssigkeit, Flexibilität und Originalität; die Jungen der Gruppe 2 waren nur in figuraler Originalität besser als die Mädchen
Stubbings 1968	5o Schüler zwischen 14,5 und 15,5 Jahren	Die nach den Ergebnissen der TTCT wenig kreative Gruppe (die unteren 27 Prozent) enthielt mehr Mädchen als Jungen, in der hochkreativen Gruppe unterschieden sich die Anzahlen von Jungen und Mädchen nicht
Solomon 1968	722 Schüler 1., 3. und 5. Klassen	Sämtliche 7 von Torrance empfohlenen Kreativitätsmaße - 3 verbale und 4 figurale Gesamtwerte - wurden aufgrund aller Torrance-Aufgaben ermittelt. Wenn signifikante Unterschiede zwischen Jungen und Mädchen auftraten, waren sie zugunsten der Mädchen
Mc Cutcheon 197o	1oo Viertkläßler	Die Mädchen hatten in allen 7 Kreativitätsmaßen außer in figuraler Originalität höhere Mittelwerte als Jungen, der Unterschied war aber nur für verbale Flexibilität und figurale Elaboration signifikant
Dewing 197o a	394 Siebtkläßler	Jungen und Mädchen waren in 3 verschiedenen Leistungsgruppen etwa in gleicher Anzahl vertreten; die Gruppen wurden aufgrund des Gesamtwertes aus Flüssigkeits- und Originalitätsscores für 4 verbale und figurale Torrence-Aufgaben gebildet
Ogletree 1971	1165 Schüler 3.- 6. Klassen aus öffentlichen und Waldorfschulen in England, Deutschland und Schottland	Mädchen waren signifikant besser als Jungen in allen verbalen und figuralen Maßen der TTCT, nur die schottischen Jungen waren in den figuralen Maßen insignifikant besser als die Mädchen
Hatcher, Felker und Treffiner 1974	188 Schüler 4.-6.Klassen	Keine Geschlechtsunterschiede in den verbalen Flüssigkeits-, Flexibilitäts- und Originalitätswerten der TTCT
Seiffge-Krenke 1974	88 Viertkläßler	Kein Geschlechtsunterschied in 27 von 28 Kreativitätsmaßen (4 Maße für jede von 4 Aufgaben, 4 verbale und 4 figurale Gesamtwerte und 4 Gesamtwerte aus verbalen und figuralen Verfahren); nur für verbale Elaboration ergab sich eine signifikante Differenz (für wen, wird nicht mitgeteilt, vermutlich für Mädchen)
Phillips und Hudgins 1974	6o Studenten	Kein Geschlechtsunterschied in dem Originalitäts- und Elaborationswert der Aufgabe "Circles"

In 9 von 18 Untersuchungen wurde kein geschlechtsspezifischer Unterschied in den TTCT-Ergebnissen festgestellt. Dabei handelte es sich um Untersuchungen auf den verschiedensten Altersstufen - von der Vorschulzeit bis zum Beginn des Erwachsenenalters. 7 der 9 Untersuchungen verwendeten Gesamtwerte aus allen oder mehreren Aufgaben der TTCT.

Eine Überlegenheit männlicher Probanden wurde in 3 Untersuchungen gefunden,deren Stichproben ebenfalls einen größeren Altersbereich abdecken. Eine der Untersuchungen verwendete nur eine Torrance-Aufgabe, eine macht keine diesbezüglichen Angaben, das Ergebnis der dritten basiert auf einer ungewöhnlich geringen Anzahl von Probanden (27 Prozent von 5o!).
Bessere Ergebnisse für Mädchen wurden in 4 Untersuchungen - für Schüler unterer und mittlerer Klassenstufen - gefunden; in einer dieser Untersuchungen wurde nur eine einzige Aufgabe aus den TTCT durchgeführt. Die Resultate von Solomon (1968) und von Ogletree (1971) basieren auf großen bzw. besonders großen Probandenzahlen und auf den verbalen und figuralen Gesamtwerten aller Torrance-Aufgaben.

Zwei Untersuchungen fanden z.T. bessere Ergebnisse für Jungen, z.T. waren die Mädchen überlegen. In einer dieser Untersuchungen wurde nur je eine verbale und eine figurale Aufgabe durchgeführt, die Zahl der Probanden war relativ klein. Die Ergebnisse der anderen Untersuchungen beruhen ebenfalls auf den Kreativitätsmaßen einzelner Aufgaben.

Insgesamt betrachtet können die von Torrance vermuteten geschlechtsspezifischen Unterschiede in den TTCT-Ergebnissen aufgrund dieser Zusammenstellung von Untersuchungsresultaten keineswegs als gesichert betrachtet werden. Vor allem bei der Verwendung verbaler und figuraler Gesamtwerte oder gar des Gesamtwertes aus verbalen und figuralen Maßen war die für die Einzelaufgaben häufiger berichtete Geschlechtsabhängigkeit der Ergebnisse meist nicht feststellbar.
Nach den Ergebnissen von Ogletree (1971) und von Solomon (1968) besteht die Möglichkeit einer leichten Überlegenheit der Mädchen, die eventuell durch ihre größere Anstrengungsbereitschaft in der schulähnlichen Testsituation zu erklären ist.

Wenn sich die Testergebnisse von Jungen und Mädchen unterschieden, waren letztere eher in verbalen Maßen besser als in figuralen und

eher in Flüssigkeit und Elaboration, während Jungen häufiger in figuralen Maßen und in der Auswertungskategorie Originalität überlegen waren.

Entwicklungsspezifische Trends sind aus Tabelle 2 nicht ersichtlich. Die Annahme, daß sich zu Beginn der Forschungsarbeit mit den TTCT die Geschlechtsrollenstereotype stärker zugunsten der Kreativitätstestergebnisse von Jungen auswirkten, ihr Einfluß im Lauf der Zeit jedoch zurückgegangen ist[1], läßt sich aus den nach dem Veröffentlichungszeitpunkt geordneten Ergebnissen der Tabelle 2 nicht bestätigen. Um die eventuellen Effekte eines solchen nicht gerade dramatisch verlaufenden kulturellen Wandels aufzuzeigen, bedarf es vermutlich systematischer Untersuchungen von größerem Umfang als die hier zusammengestellten Arbeiten.

Die Möglichkeit einer eventuellen Überlegenheit der Mädchen scheint nicht nur für die TTCT, sondern für alle in der Schule durchgeführten Kreativitätstests zu bestehen, da Mädchen auf schulische oder als schulisch wahrgenommene Anforderungen im allgemeinen mit größerer Anstrengungsbereitschaft reagieren. Da der Test-Überlegenheit jedoch keineswegs eine Überlegenheit in realen kreativen Leistungen entspricht, müßte bei der Konstruktion von Kreativitätstests ein eventueller Bias zugunsten der Mädchen ausgeglichen werden.

2.234 Schichtabhängigkeit der TTCT-Ergebnisse

In Untersuchungen an Personen, deren Kreativität aufgrund gesellschaftlich anerkannter Leistungen festgestellt wurde, wurden meist Hinweise auf den höheren Bildungsstand und die höhere berufliche Position der Herkunftsfamilien gefunden. Oden (1968, zitiert nach Freeman, Butcher und Christie 1968), die einen Teil der von Terman untersuchten hochbegabten Kinder über 4o Jahre weiterverfolgte, verglich die im Hinblick auf Berufserfolg und kreative Leistungen 1oo erfolgreichsten mit den 1oo nach diesen Kriterien erfolglosesten Individuen, die aus einer Gruppe von 8oo Hochbegabten ausgelesen worden waren. Die 1oo "Erfolgreichen" kamen überproportional häufig aus Familien, in denen Väter und Großväter College-Ausbildung

1) Die Hypothese beruht auf einer Generalisierung von Ergebnissen von Rosenberg und Sutton-Smith (196o, zitiert nach Janis 1968), nach denen die Mädchen 196o viel mehr "männliche Spiele" bevorzugten als 193o.

hatten, in denen Bücher in großer Anzahl vorhanden waren und in denen die Väter geachtete Gemeindemitglieder waren.
Auch Mac Kinnon (1965 b) und Roe (1953, beide zitiert nach Freeman, Butcher und Christie 1968) weisen darauf hin, daß in den Familien ihrer Kreativen erfolgreiche und geachtete Verwandte als Identifikationsmodelle zur Verfügung standen. Schaefer und Anastasi (1968) berichten über den Familienhintergrund ihrer nach mehreren Kriterien als kreativ eingestuften High-School-Jungen,daß es sich meist um Akademikerfamilien handelte, daß die Eltern im Hinblick auf bestimmte Interessen häufig als Vorbilder dienten und daß die bevorzugte Freizeitaktivität der Väter Lesen war. Nach Drevdahl (1964) übte die Familie seiner Gruppe kreativer Psychologen einen stärkeren Druck in Richtung auf gute Schulleistungen aus, als das bei zwei nichtkreativen Kontrollgruppen der Fall war. Diese Angaben weisen direkt oder indirekt darauf hin, daß kreative Persönlichkeiten überwiegend aus Mittel- oder Oberschichtfamilien stammen. Daraus ergeben sich für die Kreativitätsmessung und die Vorhersage kreativer Leistungen durch Tests einige Probleme und Fragen. Als erstes sollen die Fragen erörtert werden,ob die TTCT-Ergebnisse von der Sozialschicht der Probanden beeinflußt werden und - für den Fall, daß dies zutrifft - welche Sozialschicht durch diesen Kreativitätstest benachteiligt wird.

Die letztgenannte Frage erscheint zunächst überflüssig, da für eine Reihe von Intelligenztests, vor allem für viele Gruppenverfahren, eine Benachteiligung der unteren Sozialschichten festgestellt wurde.[1)] Da mit den TTCT ebenfalls überwiegend Gruppentestungen durchgeführt wurden und da überdies die Ergebnisse dieses Kreativitätstests mit Intelligenztestergebnissen signifikant korrelieren, scheint eine Benachteiligung der unteren Sozialschichten auf der Hand zu liegen.

Torrance (1971b)nennt jedoch 5 Untersuchungen mit jüngeren und älteren Schülern, in denen alle oder ein Teil der aus den figuralen Aufgaben der TTCT abgeleiteten Maße die Probanden aus den höheren Sozialschichten benachteiligten; für die verbalen Maße war es entweder umgekehrt oder die Ergebnisse für die beiden aufgrund der Schichtzugehörigkeit gebildeten Gruppen unterschieden sich nicht signifikant.

1) S. dazu Horn 1974

In 2 weiteren Untersuchungen wurden keine signifikanten schichtspezifischen Unterschiede gefunden.
Das überraschende Resultat der höheren Testergebnisse unterprivilegierter Kinder[1] in den figuralen Maßen der TTCT erklärt Torrance damit, daß der Mangel an teurem Spielzeug die Kinder dazu zwinge, mit allen möglichen Materialien zu improvisieren und daß die Familiengröße und der Lebensstil benachteiligter Familien zur Entwicklung von Problemlösungsfähigkeiten und zu Gruppenaktivitäten führen würden. Torrance scheint bei dieser Erklärung vor allem an die schwarze Bevölkerung der USA gedacht zu haben, denn er behauptet, in unterprivilegierten Familien würde außerdem mehr Wert auf Musik, Rhythmus, Tanz, mimische und gestische Ausdrucksfähigkeit und Humor gelegt als in Familien mit höherem sozioökonomischem Status. Ob die geschilderten Vorteile der Kinder sozial benachteiligter Familien deren kulturelle Deprivation im Hinblick auf die Entwicklung kreativer Fähigkeiten nicht nur ausgleichen, sondern ihnen sogar einen Vorsprung verschaffen, erscheint sehr fraglich.
Anhand einiger weiterer Arbeiten zu diesem Thema soll überprüft werden, ob die von Torrance ausgewählten Untersuchungen repräsentativ sind.
O'Bryan und Mac Arthur (1969) fanden keine Beziehung zwischen dem sozioökonomischen Status und einem figuralen TTCT-Gesamtwert; Probanden waren 85 neunjährige Jungen.
Keenan (197o) konnte für 34o Sechstkläßler ebenfalls keinen Zusammenhang zwischen dem sozioökonomischen Status und einem aus dem figuralen Gesamtscore der TTCT und 2 weiteren Kreativitätstests kombinierten Kreativitäts-Gesamtwert feststellen.

Lytton und Cotton (1969) fanden eine positive Korrelation zwischen Vaterberuf und einem Gesamtwert für 5 Aufgaben der TTCT von .26. Sie weisen darauf hin, daß diese Beziehung weniger eng ist als die zwischen Vaterberuf und der Fähigkeit zu konvergentem Denken (gemessen mit dem "Verbal Reasoning Quotient"), die für ihre Stichprobe von 143 14jährigen Schülern .58 betrug.
Dewing (197o b) ließ die 1o Schulen, in deren 7. Klassen 394 Schüler mit 2 verbalen und 2 figuralen Torrance-Aufgaben untersucht worden waren, durch 3 unabhängige Beurteiler nach dem sozioökonomischen Niveau in eine Rangreihe bringen (die Reliabilität der Ratings wird mit .93 angegeben) und korrelierte die Mittelwerte der TTCT-

---

1) Die Vermutung, daß diese Probandengruppen unterprivilegiert waren, wird aus der häufigen Verwendung der Kennzeichnung dieser Gruppen als "disadvantaged" hergeleitet.

Gesamtwerte aus verbalen und figuralen Maßen für jede Schule mit ihrer sozioökonomischen Rangposition; der Rangkorrelationskoeffizient war .60.

Die umfassendste Untersuchung wurde von Ogletree und Ujlaki (1973) in englischen, schottischen und deutschen Elementarschulen (6 staatliche und 6 Waldorf-Schulen) an 1165 Schülern durchgeführt. Sie teilten die Probanden nach dem Vaterberuf in 3 sozioökonomische Gruppen ein, in "upper-middle","middle" und "lower". Die sozioökonomisch höchste Gruppe hatte in allen 8 Gesamtmaßen aus sämtlichen TTCT-Aufgaben signifikant höhere Werte als die beiden anderen Gruppen außer in figuraler Flexibilität, für die der Unterschied nicht signifikant war. Der Unterschied zwischen mittlerer und unterer Gruppe war etwas weniger ausgeprägt, d.h. für einen Teil der Ergebnisse - verbale Flüssigkeit, figurale Flüssigkeit, Flexbilität, Elaboration und den figuralen Gesamtwert - war der Unterschied zugunsten der mittleren Gruppe nicht signifikant. Für Jungen und Mädchen getrennt durchgeführte Analysen erbrachten nur für die verbalen Testwerte signifikant bessere Ergebnisse der jeweils höheren Gruppen; in den figuralen Maßen erzielte die sozioökonomisch höchste Gruppe bei Jungen und Mädchen ebenfalls die höchsten Werte, aber die Differenz war nicht signifikant.
Somit zeigte sich auch hier die stärkere Schichtabhängigkeit der verbalen Maße, aber die von Torrance vermutete Überlegenheit der Kinder unterer Sozialschichten in den figuralen Maßen wurde nicht bestätigt.
Wenngleich durch den hohen Anteil an Waldorf-Schulen die höchste und mittlere Gruppe stark überrepräsentiert waren, lassen die Ergebnisse von Ogletree und Ujlaki es doch äußerst unwahrscheinlich erscheinen, daß unterprivilegierte Kinder in figuralen Kreativitätsaufgaben ihren sozioökonomisch besser gestellten Altersgenossen überlegen sind. Erstere hatten in sämtlichen figuralen Maßen schlechtere Ergebnisse, die Unterschiede waren lediglich z.T. nicht signifikant.

Die Gesamtheit der hier dargestellten Ergebnisse zur Frage der Schichtabhängigkeit der TTCT läßt es angezeigt erscheinen, bei Untersuchungen auf dem Gebiet der Kreativitätsmessung die Möglichkeit einer Beeinflussung der Testergebnisse durch den sozioökonomischen Status der Probanden zu überprüfen.

Falls sich schichtspezifische Unterschiede zeigen, sehen sich Testkonstrukteure und Untersucher, die kreatives Verhalten und kreative Leistungen vorhersagen wollen, vor die Frage gestellt, ob schichtspezifische Normen bereitgestellt bzw. ob bei vorliegenden Testergebnissen ggf. eine Anhebung der Testwerte benachteiligter Gruppen vorgenommen werden soll.
Da nach den eingangs erwähnten Untersuchungen von Drevdahl, Mac Kinnon u.a. viele aufgrund kreativer Leistungen bekannte Individuen aus Mittel- und Oberschicht-Familien kommen und dieses Resultat sehr gut als Auswirkung schichtspezifischer Erziehungsziele und Eltern-Kind-Interaktionen erklärt werden kann,[1] hätte eine schichtspezifische Normierung bzw. eine Anhebung der Testergebnisse der Probanden aus unteren Schichten vermutlich eine Verminderung der prädiktiven Validität des Kreativitätstests zur Folge. Wird dagegen ein schichtspezifischer Bias eines Tests nicht korrigiert, kann dies zu einer Konservierung der Unterschiede führen, weil bei Entscheidungen über Schullaufbahn, Förderungsprogramme u.ä. Kinder aus unteren Sozialschichten aufgrund der für sie ungünstigeren Testergebnisse benachteiligt werden. Vermutlich kann diese Frage nicht generell entschieden werden, sondern muß unter Berücksichtigung des jeweiligen Untersuchungsziels beantwortet werden. Bei Vorhersage-Untersuchungen muß vor allem gefragt werden, ob langfristige Fördermaßnahmen zu erwarten sind oder jedenfalls im Bereich des Möglichen liegen oder nicht. Auf keinen Fall kann jedoch davon ausgegangen werden, daß die Ergebnisse von Kreativitätstests im Gegensatz zu denjenigen von Intelligenz- und Schulleistungstests von der Sozialschicht der Probanden nicht beeinflußt würden.

2.235 Zusammenhänge zwischen TTCT-Ergebnissen, Intelligenz und Schulleistung

Die Frage des Zusammenhangs zwischen den Ergebnissen des Torrance-Tests und Intelligenztestwerten wurde ebenso wie diejenige nach der Beziehung zwischen TTCT-Resultaten und Schulleistungsmaßen bereits

1) S. dazu - abgesehen von den erwähnten Veröffentlichungen - Christie (1970), Dewing (1970c), Dreyer und Wells (1966, zitiert nach Dewing 1970 c), Domino (1969), Mollenhauer (1971[6]), Nichols (1964), Roe (1963, zitiert nach Datta und Parloff 1967), Weisberg und Springer (1961)

mehrmals angeschnitten. Es wurde aufgezeigt, daß die Kreativitäts-Intelligenz-Korrelation höher ist als von Torrance vermutet (vgl. S.2o-22 und S. 45-49) und daß die Zusammenhänge zwischen TTCT-Ergebnissen und Schulleistungen überwiegend positiv, teilweise signifikant, aber fast immer niedrig sind (vgl. S. 34-36).

Gemeinsam war den Untersuchungen zu beiden Fragen, daß die Ergebnisse sehr unterschiedlich und z.T. verwirrend und widersprüchlich waren.
Torrance (1962a)versuchte, diese Unstimmigkeiten - abgesehen von dem Hinweis auf schulspezifische Variablen - durch seine "Schwellenhypothese der Intelligenz" zu erklären, nach der kreative Fähigkeiten erst ab einem Intelligenzquotienten um 12o für die Schulleistungen wichtig werden, während eine weitere Erhöhung des IQ nicht zur Erhöhung der Schulleistungen führt.
Da die Beziehung zwischen Kreativität, Intelligenz und Schulleistungen für Untersuchungen zur Kreativität von Schulkindern immer von Interesse sein dürfte, soll an einigen Untersuchungen, die unselegierte Stichproben verwendeten, geprüft werden, ob die empirische Evidenz für die Hypothese von Torrance spricht oder ob sich eine andere Form der Beziehung zwischen den 3 Variablen abzeichnet.
Daß der Zusammenhang zwischen Intelligenztest- und TTCT-Ergebnissen höher ist als Torrance angenommen hatte, und daß die Korrelationskoeffizienten bei Einbeziehung des gesamten Streubereichs der Intelligenz je nach den verwendeten Meßinstrumenten zwischen .13 und .54 liegen, kann aus Tabelle 1 entnommen werden.[1)]
Die Hypothese, daß die Kreativitäts-Intelligenz-Korrelation mit steigendem IQ abnimmt, wurde durch die Ergebnisse von Haddon und Lytton (1968) für 211 elf- bis zwölfjährige Schüler und durch diejenigen von Yamamoto (1965b)für 461 und 827 Fünftkläßler gestützt. Sie wurde von Guilford und Hoepfner (1966, zitiert nach Guilford und Christensen 1973) und von Guilford und Christensen (1973) unter Verwendung von Guilford-Aufgaben bestätigt und durch den Nachweis dreieckiger Streuungsdiagramme veranschaulicht (vgl. S. 8). Obwohl auch Butcher (1973) Belege für die Richtigkeit dieser Hypothese

---

1) S.S.47-48. Für die Angaben des Streubereichs der Intelligenz-Kreativitäts-Beziehung wurden nur verbale und figurale Kreativitäts-Gesamtmaße sowie Gesamtwerte aus verbalen und figuralen Maßen berücksichtigt.

anführt (Nuttal 1971 und Ogilvie 197o), ist das Bild nicht widerspruchsfrei: Bennett (1973), der 331 Zehnjährige mit semantischen Guilford-Aufgaben und dem Verbal Reasoning Quotient untersuchte, fand auch auf hohem Intelligenzniveau eine relativ enge Beziehung zwischen Intelligenz und Kreativität und Dacey und Madaus (1971) fanden nur für die größte von 3 Stichproben (N_1=867, N_2=583 und N_3=182) eine signifikante Kreativitäts-Intelligenz-Korrelation und die erwartete Abnahme des Zusammenhangs mit steigendem IQ.
Insgesamt gesehen spricht allerdings die Mehrzahl der Untersuchungen dafür, daß die Kovariation von Intelligenz und Kreativität auf niedrigen und mittleren Intelligenzstufen größer ist als auf höheren. Nach der Schwellenhypothese von Torrance könnte vermutet werden, daß die weitgehende Unabhängigkeit der beiden Variablen relativ abrupt bei einem IQ um 12o beginnt (womit erklärt wäre, warum die Kreativitätstestergebnisse bei Kindern dieses Intelligenzniveaus zur Vorhersage der Schulleistungen beitragen, bei weniger intelligenten jedoch nicht), dafür läßt sich jedoch weder in den Streuungsdiagrammen Guilfords et al.[1] noch in den Ergebnissen von Bennett (1973), Dacey und Madaus (1971), Haddon und Lytton (1968) und von Yamamoto (1965b)ein Hinweis finden. Die Verringerung der Kreativitäts-Intelligenz-Korrelation erscheint vielmehr allmählich und tritt in manchen Untersuchungen schon auf wesentlich niedrigerem Intelligenzniveau in Erscheinung; bei Haddon und Lytton ist der Koeffizient z.B. für Probanden mit einem Verbal Reasoning Quotienten unter 1oo .512, bei solchen mit einem VRQ von 1oo und mehr .164.
Im Hinblick auf den Zusammenhang zwischen Kreativität und Schulleistung wäre nach der Schwellenhypothese keine oder eine sehr niedrige Korrelation für Schüler mit einem IQ unter 12o und eine wesentlich engere und auf jeden Fall signifikante Beziehung für Schüler mit höherem IQ zu erwarten.[2] Diese Annahme wurde auf mehrere Arten überprüft:
Edwards und Tyler (1965) stellten aus einer größeren Gruppe von Schülern neunter Klassen eine Gruppe von 36 Probanden mit hohen Ergebnissen im School and College Ability Test (oberes Drittel) und durch-

1) S. dazu Guilford und Christensen 1973

2) Daß bei Berücksichtigung des gesamten IQ-Bereichs der Zusammenhang zwischen Kreativität und Schulleistung bedeutungslos ist, wurde bereits bei der Besprechung der Übereinstimmungsvalidität dargestellt (vgl. S. 34-36).

schnittlichen Gesamtwerten aus 2 TTCT-Aufgaben und eine zweite Gruppe von 26 Schülern mit hohen SCAT- und TTCT-Ergebnissen (oberes Drittel für beide Variablen) zusammen. Beim Vergleich der Schulleistungstestergebnisse zeigte sich keine Überlegenheit der letztgenannten Gruppe, obwohl die auf dem SCAT basierende IQ-Schätzung für beide Gruppen bei 123 lag und somit nach der Schwellenhypothese eine bessere Schulleistung der intelligenten und kreativen Gruppe zu erwarten war. (Ein Vergleich der Noten beider Gruppen erbrachte das interessante Resultat, daß diese Gruppe einen signifikant niedrigeren GPA hatte als die intelligente, aber weniger kreative Gruppe.) Yamamoto (1967a) verglich die TTCT-Mittelwerte von Schülern mit erwartungswidrig guten und erwartungswidrig schlechten Schulleistungen (die Erwartungen basieren auf dem IQ der Schüler). Er fand keine Anhaltspunkte dafür, daß Overachiever kreativer sind als Underachiever.

Cicirelli (1965) untersuchte die Schwellenhypothese und die Möglichkeit irgendwelcher Interaktionen von Intelligenz und Kreativität in ihrer Auswirkung auf die Schulleistung an 609 Sechstkläßlern mit Hilfe einer Varianzanalyse. Er schließt aus seinen Ergebnissen: "While the relationship of creativity and achievement was a weak one, the form of the relationship was such that IQ and creativity were additive and linear in their effect on academic achievement." (S. 308)

Yamamoto und Chimbidis (1966) teilten ihre Stichprobe von 790 Fünftkläßlern in 2 Gruppen auf, in diejenigen mit einem IQ bis 120 (N=604) und diejenigen mit einem IQ über 120 (N=186). In der ersten Gruppe lag die Kreativitäts-Intelligenz-Korrelation bei .26, in der zweiten bei .01. Die Beziehung zwischen Intelligenz und verschiedenen Schulleistungsmaßen war in der weniger intelligenten Gruppe durchweg signifikant enger als in der hochintelligenten Gruppe, die Partialkorrelationen zwischen Kreativität und Schulleistung unterschieden sich jedoch nicht signifikant. Sie waren überwiegend nicht signifikant und sämtlich niedrig ($.00 \leq r \leq .20$, Md=.105). Yamamoto und Chimbidis weisen darauf hin, daß die Vorhersage der Schulleistungen durch Intelligenztestergebnisse für hochintelligente Schüler wegen des geringen Zusammenhangs kaum möglich ist und daß deshalb der relative Beitrag der Kreativitätsmaße zur Schulleistungsvorhersage für diese Gruppe größer war. Trotzdem war auch in der Gesamt-

gruppe der hochintelligenten Schüler der IQ der bessere Prädiktor der Schulleistungen als die TTCT-Ergebnisse; dies gilt in stärkerem Maße für die Mädchen als für die Jungen, deren Schulleistungen auf diesem Intelligenzniveau durch die verwendeten Prädiktoren kaum mehr vorhersagbar erscheinen.

Somit können folgende Ergebnisse als relativ gesichert betrachtet werden: Die Kreativitäts-Intelligenz-Beziehung nimmt mit steigendem IQ ab.
Deutliche Hinweise auf eine Intelligenzschwelle wurden nicht gefunden, die Korrelation zwischen Kreativität und Intelligenz scheint vielmehr mit zunehmendem IQ kontinuierlich abzunehmen.
Der Zusammenhang zwischen Intelligenz und Schulleistung ist für unselegierte Stichproben relativ eng - Korrelationen zwischen .5o und .6o dürften überwiegen -, sinkt aber auf höherem Intelligenzniveau ebenfalls stark ab (für Jungen vermutlich stärker als für Mädchen).
Die Beziehung zwischen Kreativität und Schulleistung ist sehr schwach und erreicht auch auf höherem Intelligenzniveau nicht das für Vorhersagen erforderliche Ausmaß.

2.24 Schlußfolgerungen für die weitere Arbeit

Eine Adaptation der TTCT erscheint aus mehreren Gründen nicht sinnvoll:

- Das Verfahren wurde nicht auf der Grundlage einer einheitlichen Kreativitätstheorie entwickelt.
- Weder die zahlreichen Untersuchungen zur Übereinstimmungs- noch die wenigen Untersuchungen zur prädiktiven Validität erbrachten Ergebnisse, die als Nachweis der Gültigkeit der TTCT gewertet werden könnten.
- Diejenigen Arbeiten, die von Torrance und von anderen Autoren mit dem Ziel der Konstruktvalidierung durchgeführt wurden, führten ebenfalls nicht zu Resultaten, die als Beleg für die Validität des Verfahrens herangezogen werden könnten.

Die zahlreichen, mit den TTCT durchgeführten Untersuchungen erbrachten jedoch einige Ergebnisse, die bei der weiteren Arbeit berücksichtigt werden sollen:

- Die Auswertung derselben Testaufgabe nach mehreren Kriterien ist vermutlich nicht erforderlich. Zumindest liegen keine Anhaltspunkte dafür vor, daß die Elaborationswerte valide Indikatoren kreativer Fähigkeiten sind und daß die Flexibilitätswerte Informationen erbringen, die in den Flüssigkeitsscores nicht enthalten sind.
- Die Möglichkeit geschlechts- und schichtspezifischer Unterschiede in den Testergebnissen eines Verfahrens zur Erfassung kognitiver Voraussetzungen für kreatives Verhalten muß überprüft werden.
- Bei Stichproben mit normaler Streuung der Intelligenztestwerte muß wahrscheinlich mit einer signifikanten Kreativitäts-Intelligenz-Korrelation gerechnet werden.
- Schulleistungen sind vermutlich keine geeigneten Validitätskriterien für einen Kreativitätstest.

2.3 Der Kreativitätstest von Wallach und Kogan

Bereits wenige Jahre nach Erscheinen der ersten Versionen der Torrance-Tests veröffentlichten Wallach und Kogan (1965) ihr Buch "Modes of Thinking in Young Children", in dem sie - nach einer sehr kritischen Auseinandersetzung mit den zu diesem Zeitpunkt vorliegenden Kreativitätstests - ihr eigenes Verfahren vorstellen.
Die Besprechung des Tests von Wallach und Kogan wird die gleichen Punkte berücksichtigen wie diejenigen der TTCT, wenn auch z.T. mit unterschiedlicher Gewichtung.

2.31 Theoretische Fundierung

Wallach und Kogan fassen ihre Kritik an den vorliegenden Verfahren zur Kreativitätsmessung folgendermaßen zusammen: "The measures that have been construed as indicators of creativity are not indicators of some single psychological dimension parallel to and distinct from the dimension of general intelligence defined by conventional intelligence test indices... what is needed at this point is a new attempt at conceptual analysis on which to base measurement procedures." (S. 12-13)
In Entsprechung dieser Forderung entwickeln sie ihr Meßverfahren auf der Grundlage einer einheitlichen Kreativitätskonzeption, derjenigen von Mednick (1962).

Mednick definiert kreatives Denken als "the forming of associative elements into new combinations which either meet specified requirements or are in some way useful. The more mutually remote the elements of the new combination, the more creative the process or the solution.". (S. 221)
Ausgehend von dieser Definition und einer Reihe von Untersuchungsergebnissen der Assoziationsforschung stellt er mehrere Hypothesen über die Unterschiede in den Assoziationsverläufen hochkreativer und wenig kreativer Individuen auf, die in ihrer Gesamtheit die Kreativitätstheorie bilden, die Mednick zur Grundlage seines eigenen und Wallach und Kogan zur Grundlage ihres Kreativitätstests machten. Die wichtigsten dieser Hypothesen sollen kurz dargestellt werden:

1. Je größer die Anzahl der Assoziationen, über die ein Individuum verfügt, desto größer ist die Wahrscheinlichkeit einer kreativen Lösung.

2. Die Organisation der verfügbaren Assoziationen - die Assoziationshierarchie - beeinflußt die Wahrscheinlichkeit einer kreativen Lösung und die Schnelligkeit, mit der sie auftritt.
Mednick beschreibt einen nach seiner Theorie charakteristischen Unterschied in den Assoziationsverläufen kreativer und unkreativer Individuen, der durch ihre unterschiedlichen Assoziationshierarchien bedingt ist: Letztere antworten auf ein Reizwort rasch mit einigen wenigen und üblichen Assoziationen, danach sinkt die anfänglich hohe assoziative Stärke des Reizwortes zu den assoziativen Elementen rasch ab; wenig kreative Individuen sind also durch eine "steile Assoziationshierarchie" gekennzeichnet, was bedeutet,daß sie relativ schnell einige stereotype Assoziationen liefern, ihr Reservoir aber bald erschöpft ist. Kreative dagegen haben weniger prädominante Assoziationen bzw. eine "flache Assoziationshierarchie", was bedeutet, daß sie auf ein Reizwort langsamer, stetiger und mit einer insgesamt höheren Anzahl von Assoziationen reagieren, die überwiegend weniger üblich und stereotyp sind.

3. Je häufiger ein Individuum ein bestimmtes Problem auf die gleiche Art und Weise gelöst hat, desto geringer ist die Wahrscheinlichkeit, daß er für dieses Problem - bei Verwendung derselben Materialien oder Hilfsmittel - eine kreative Lösung findet.

4. Bei dem Versuch, ein Problem zu lösen, treten meist viele Assoziationen auf. Die Auswahl der kreativen Lösung erfolgt aufgrund spezifischer Erfordernisse, die das Problem explizit oder implizit beinhaltet. Die Konsequenzen jeder assoziativen Kombination bzw. jeder eventuellen Lösung werden mit diesen Erfordernissen des Problems verglichen und diejenige Kombination, die den Erfordernissen am besten entspricht, wird ausgewählt.

5. Gelernte oder hereditär determinierte Methoden, ein Problem anzugehen, beeinflussen ebenfalls die Wahrscheinlichkeit einer kreativen Lösung. Je nach Art des Problems gelangen z.B. Individuen mit dem kognitiven Stil der "visualizer" oder der "verbalizer" eher ans Ziel.

Wallach und Kogan stellen Mednicks Hypothesen über die Assoziationsverläufe kreativer und weniger kreativer Individuen in den Mittelpunkt ihrer Kreativitätskonzeption. Nach ihrer Ansicht können sowohl Hochkreative als auch wenig Kreative zunächst mit üblichen Assoziationen reagieren, aber während letztere, die sie wie Mednick durch einen steilen Gradienten der Assoziationsstärke charakterisieren, damit ihr Assoziationsreservoir erschöpft haben, fahren erstere in der Produktion von Antworten fort, und diese Antworten entfernen sich immer mehr vom üblichen und werden immer einzigartiger (s. Abb.1). Ihr Gradient ist flacher und schneidet die Ordinate an einem niedrigeren Punkt als derjenige der wenig Kreativen, was bedeutet, daß sie langsamer antworten (Latenzzeit als Operationalisierung der Assoziationsstärke!) und berührt die Abszisse erst sehr viel später, was impliziert, daß die Hochkreativen seltene und einzigartige Assoziationen produzieren.

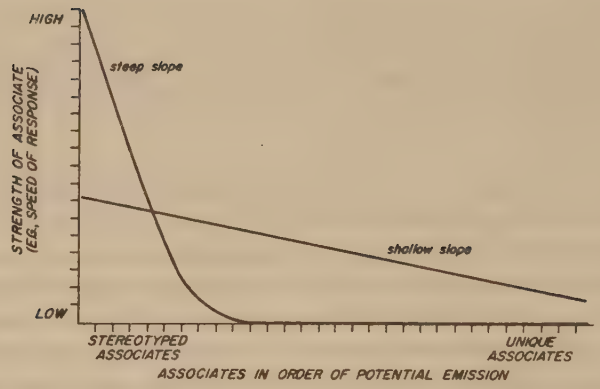


Abb. 1: Hypothetische Assoziationsverläufe kreativer und wenig kreativer Individuen (Aus Wallach, M.A. und Kogan, N.: Modes of Thinking in Young Children. New York: Holt, Rinehart and Winston 1965, S. 16)

Diese Kreativitätskonzeption führte zu mehreren Konsequenzen für die Konstruktion des Meßverfahrens:

1. Die Aufgabenstimuli müssen eine möglichst große Zahl von Assoziationen hervorrufen, zugleich müssen die Aufgaben den Probanden jedoch gewisse Beschränkungen auferlegen: Wallach und Kogan sprechen von "task constraints, so to structure a situation that only associations which are appropriate - which meet certain requirements - are given." (S. 14)[1)]

---

1) Somit handelt es sich bei dem Test von Wallach und Kogan um ein Verfahren, das sich nach Inhalt und Durchführung kaum von der Technik der kontrollierten Assoziationen unterscheidet - nicht der freien Assoziationen, wie Krause (1972, S. 73) behauptet.

2. Unter der Bedingung, daß die Assoziationen der Probanden einem Kriterium der Richtigkeit oder Angemessenheit genügen müssen, können individuelle Unterschiede in der Kreativität am besten mit Hilfe zweier Variablen erfaßt werden: der Anzahl und der Einzigartigkeit der Assoziationen.
Wallach und Kogan betrachten diese beiden Variablen nicht als unabhängig voneinander, sondern erwarten vielmehr, daß zu Beginn des Assoziationsverlaufs übliche und damit häufig genannte Assoziationen auftreten und später die selteneren und einzigartigen. Die kreativen Individuen würden also mehr Assoziationen produzieren als die weniger kreativen, und sie würden eine größere Zahl nur einmal oder jedenfalls sehr selten auftretender Assoziationen nennen (vgl. Abb. 1).

3. Einzigartige Assoziationen werden von unkreativen Individuen gar nicht und von kreativen Individuen erst in einer späteren Phase des Assoziationsverlaufs und auch dann mit langsamer Produktionsrate gegeben (vgl. Abb. 1). Daraus folgt, daß kreative Probanden durch Tests mit Zeitbegrenzungen stark benachteiligt würden; je nach Kürze der Testzeit würden sie im Hinblick auf die Anzahl der Antworten eventuell sogar schlechter abschneiden als die weniger kreativen Individuen. Probanden mit einem steilen Gradienten des Assoziationsverlaufs würden dagegen bei sehr kurzen Zeitgrenzen herausragende Leistungen in der Produktivität zeigen.
Wallach und Kogan vermuten eine inverse Beziehung zwischen der Anzahl der Assoziationen und ihrer Produktionsrate[1)] und folgern daraus, daß ein Kreativitätstest keine Zeitbegrenzung haben darf.

4. Durch Mednicks Aussagen über eventuelle Auswirkungen unterschiedlicher kognitiver Stile auf die assoziative Produktivität (vgl. S. 79) sehen sich Wallach und Kogan veranlaßt, die Aufgabenstimuli, die zugleich die Kriterien für die Beurteilung der Angemessenheit der Antworten beinhalten, zum Teil in verbaler und zum Teil in visueller Form darzubieten.

---

1) Diese Annahme wird durch Untersuchungsergebnisse von Bousfield und Sedgewick (1944, zitiert nach Wallach und Kogan 1965) gestützt. Nach Christensen, Guilford und Wilson (1957) sind später auftretende Antworten bessere Maße von Originalität i.S. von statistischer Seltenheit und von Entlegenheit der Assoziationen ("remoteness") als die zeitlich früher auftretenden.

Nicht aus der Kreativitätskonzeption Mednicks, sondern aus introspektiven Berichten hochkreativer Persönlichkeiten abgeleitet ist ein weiteres Charakteristikum der Testsituation von Wallach und Kogan: Die Probanden sollen keineswegs das Gefühl haben, sich in einer Prüfungssituation zu befinden, damit nicht statt der aufgabenorientierten ichzentrierte Motivationen entstehen. Deshalb soll der Test in einer entspannten, spielähnlichen Atmosphäre durchgeführt werden.
Eine der Maßnahmen, durch die eine solche Atmosphäre herbeigeführt werden sollte, war die Einzeltestung durch einen den Kindern vertrauten Testleiter, eine weitere bestand darin, die Testaufgaben als Spiele zu deklarieren; günstiger als die Änderung der Bezeichnung dürfte es sich ausgewirkt haben, daß die Kinder ihre Antworten nicht selbst aufschreiben, sondern nur mündlich mitteilen mußten. Unter diesen Testbedingungen erwarteten Wallach und Kogan, daß die 5 Untertests ihres Verfahrens eine einheitliche Persönlichkeitsdimension von ähnlich hoher interner Konsistenz wie die Subtests von Instrumenten zur Messung der allgemeinen Intelligenz erfassen und daß diese Persönlichkeitsdimension sich als unabhängig von der allgemeinen Intelligenz erweist. Sie weisen darauf hin, daß die Demonstration der Unabhängigkeit ihrer assoziationstheoretisch konzipierten und operationalisierten Kreativitätsdimension von derjenigen der allgemeinen Intelligenz besondere Beachtung verdiene, weil die Ergebnisse ihres Kreativitätstests u.a. direkt vom Inhalt des Assoziationsspeichers bzw. des Repertoires an "kognitiven Elementen"[1] abhängen, während die Fähigkeit, kognitive Elemente zu speichern, als wichtige Determinante der allgemeinen Intelligenz betrachtet wird.

Dagegen ist allerdings einzuwenden, daß die Behauptung einer Abhängigkeit der Kreativitätstestergebnisse von der Anzahl der vorhandenen Assoziationen schlechthin nicht beweisbar ist, da immer nur Aussagen über die abrufbaren, nie aber über die tatsächlich vorhandenen Assoziationen möglich sind.

---

1) Nach Krause (1972) betrachtet Mednick die Assoziationen als repräsentative Stichprobe aller kognitiven Elemente, während Wallach und Kogan beide Begriffe synonym verwenden.

Dies wird von Wallach und Kogan an sich auch gesehen, wie ihre Weiterentwicklung der theoretischen Fundierung ihres Tests durch Einbeziehung der psychoanalytischen Konzepte Angst und Tendenz zur Verdrängung beweist. Diese beiden Konstrukte wurden über die Variable "Testsituation" so eng mit den assoziationstheoretischen Grundlagen des Verfahrens verknüpft, daß sie bereits an dieser Stelle und nicht - wie bei Wallach und Kogan - im Rahmen der Konstruktvalidierung **besprochen werden sollen.**

Wallach und Kogan definieren Angst in Anlehnung an Freud als einen unangenehm erlebten Bewußtseinszustand,der als Folge von inneren, die Integrität des Ich bedrohenden Gefahren auftritt. Sie führen eine ganze Reihe empirischer Belege dafür an, daß Angst und Intelligenztestwerte negativ korrelieren und daß Angst der kausale Faktor dieser Beziehung ist.
Im Hinblick auf die Ergebnisse ihres Kreativitätstests können nach Ansicht der beiden Autoren widersprüchliche Hypothesen aufgestellt werden: Wegen der entspannten, spielähnlichen Testsituation - in der z.B. hochängstliche Kinder größere Verbesserungen der Intelligenztestwerte erzielten als weniger ängstliche - erwarten sie gute Ergebnisse der Hochängstlichen. Wegen ihrer Vorsicht bei der Beantwortung der Testaufgaben könnte jedoch auch angenommen werden, daß sie nur wenige Assoziationen produzieren; andererseits wird aber auch die Möglichkeit gesehen, daß die ängstlichen Kinder die Aufgaben mit einer geradezu zwanghaften Ausdauer bearbeiten, anstatt ihrem Wunsch zu folgen und abzubrechen.

Unter der Tendenz zur Verdrängung verstehen Wallach und Kogan einen bestimmten durchgängigen"Verhaltensstil" im Umgang mit der eigenen Angst.[1] Defensive Individuen müssen für die Verdrängung angsterregender Inhalte sehr viel Energie aufbringen und sind daher in ihren Wahrnehmungs- und Handlungsmöglichkeiten stark eingeschränkt.

Der Begriff "defensiveness" wird in Anlehnung an Rogers in Gegensatz zu **"openness to experience"** gebracht, wodurch seine Beziehung zu kreativem Verhalten deutlich wird[2].

---

1) "Defensiveness" hat zwar an sich eine eingeschränktere Bedeutung i.S. von Überempfindlichkeit gegen Kritik an der eigenen Person (s. English und English 1958), aber daß dieser Begriff, "denial" und "repression" synonym verwendet werden (vgl. Wallach und Kogan 1965, S. 196), weist darauf hin, daß die beiden Autoren eher eine Abwehrbereitschaft gegen alle möglichen Bedrohungen des Ich darunter verstehen.

2) S. dazu z.B. MacKinnon 1962, 1967

Wallach und Kogan gehen davon aus, daß defensive Kinder aufgrund ihrer starken Abwehrmechanismen mit der Angst, die durch die ihnen vertrauten Intelligenztestdurchführungen hervorgerufen wird, leicht fertig werden. Sie erwarten jedoch aufgrund von Untersuchungsergebnissen, daß diese Kinder eine unbekannte und relativ unstrukturierte Testsituation, wie sie bei dem Verfahren von Wallach und Kogan vorliegt, als bedrohlich erleben und deshalb Abwehrprozesse mobilisieren, die ihre Testleistung stärker beeinträchtigen als der evaluative Kontext eines Intelligenztests. Da einerseits ein gutes Ergebnis in dem Kreativitätstest von Wallach und Kogan die Verfügbarkeit entlegener Erlebnisinhalte und die Herstellung von Beziehungen zwischen bis zu diesem Zeitpunkt unverbundenen kognitiven Elementen voraussetzt und andererseits eine hohe Abwehrtendenz die perzeptive und kognitive Offenheit gegenüber Erfahrungen und die Fähigkeit zur Ausnutzung der im Repertoire des Individuums vorhandenen Cues stark herabsetzt, stellen Wallach und Kogan die Hypothese einer inversen Beziehung zwischen Kreativität und Tendenz zur Verdrängung auf.

Das Verfahren von Wallach und Kogan dürfte der einzige Kreativitätstest sein, der aus einer einheitlichen Kreativitätstheorie entwickelt wurde. Diese Theorie ist in sich schlüssig, berücksichtigt bis dahin vorliegende wissenschaftliche Erkenntnisse - vor allem aus der Assoziationsforschung, aber auch Untersuchungsergebnisse zur Auswirkung von verschiedenen Testsituationen, von Angst und Abwehrbereitschaft auf Intelligenztestleistungen - und hat vor allem den Vorteil, daß sie die Ableitung einer ganzen Reihe empirisch prüfbarer Hypothesen und damit eine Konstruktvalidierung sowohl des Meßinstruments als auch der zugrunde liegenden Theorie ermöglicht.
Einige offene Fragen und Kritikpunkte sollen noch kurz besprochen werden.
Zunächst fällt auf, daß Wallach und Kogan keine Aussagen über die Entstehung der interindividuellen Unterschiede im Umfang des Assoziationsreservoirs und in der Form der Assoziationshierarchie machen. Mednick und Mednick (1964) weisen zwar darauf hin, daß die häufige Wiederholung derselben Antwort auf einen vorgegebenen Stimulus zu einer Hierarchie mit hoher assoziativer Stärke für diesen Reiz führt, und würden die entgegengesetzte Entwicklung eines flachen Gradienten der Assoziationsstärke vermutlich durch ihren "drive for associative novelty" (S. 61)[1] erklären, aber Wallach und

1) Mednick und Mednick (1964) bevorzugen ausdrücklich den Begriff "drive" und versuchen, seine Existenz experimentell nachzuweisen.

Kogan greifen weder auf diese Erklärung zurück noch bieten sie eine eigene Hypothese an. Dies ist umso erstaunlicher, als Krause (1972) gezeigt hat, daß durch eine Verknüpfung assoziationstheoretischer Begriffe mit psychoanalytischen Gedankengängen sehr wohl genetische Erklärungen möglich werden und Wallach und Kogan die Verbindung zwischen dem für die Genese der Unterschiede zentralen Vorgang der Verdrängung und ihrem Kreativitätskonzept sehr wohl gesehen haben (vgl. S. 83-84).
Ein zweiter Kritikpunkt betrifft die Forderung der Unabhängigkeit der Kreativitäts- von den Intelligenztestergebnissen. Der Nachweis, daß durch die beiden Verfahren zwei verschiedene Persönlichkeitsdimensionen erfaßt werden, kann auch auf andere Weise geführt werden als mit Hilfe von Nullkorrelationen zwischen ihren Ergebnissen.[1)] Vor allem aber muß stark bezweifelt werden, ob durch Kreativitätstests, die völlig unabhängig von Intelligenztestergebnissen sind, gesellschaftlich bedeutsame, kreative Leistungen vorhergesagt werden können. Nicholls (1972) zeigt die Konsequenzen der Forderung Wallachs und Kogans auf: "... if creativity measures are unrelated to intelligence, this suggests that capacity for original, culturally meaningful achievement is unrelated to intelligence. (The alternative is that creativity has no connection with the capacity for original, meaningful achievement.)" (S. 72o)

Ein weiterer Einwand richtet sich gegen die Auswertung des Kreativitätstests nach Anzahl und Originalität (i.S. von Einzigartigkeit) der Antworten bei unbeschränkter Bearbeitungszeit. Nach dem von Mednick übernommenen Modell der Assoziationsverläufe (vgl. Abb. 1, S.8o) ist Übereinstimmung oder zumindest eine sehr hohe Korrelation zwischen den beiden Kreativitätsmaßen zu vermuten. Wallach und Kogan scheinen den durch das Modell angezeigten sehr engen - bei zeitlich unbegrenzten Testungen fast perfekten - Zusammenhang nicht zu erwarten,[2)] sonst würden sie die zeitaufwendige Auswertung nach der Einzigartigkeit der Antworten nicht ins Auge fassen, aber sie

---

1) Z.B. durch eine Faktorenanalyse, mit Hilfe der "multitrait-multimethod matrix" von Campbell und Fiske (1959) oder dadurch, daß gezeigt wird, daß der Kreativitätstest bestimmte Kriterien besser vorhersagt als ein Intelligenztest und umgekehrt.

2) S. dazu Wallach und Kogan (1965, S. 14); "This is not to assert that these variables necessarily would be independent of each other; rather, it is quite possible that more frequent associations will occur earlier and more unique associations later in a sequence.." (Unterstreichung vom Verfasser).

geben keinerlei Begründung für diese nach ihrer eigenen Theorie redundante Originalitätsauswertung.
Ein letzter Kritikpunkt betrifft die Hypothese, daß Kreativität nur durch ein Meßinstrument ohne Zeitbegrenzung erfaßt werden könne. Die Ableitung dieser Annahme aus der assoziationstheoretischen Kreativitätskonzeption ist zwingend, aber Wallach und Kogan lassen jegliche Erörterung alternativer Interpretationen der Meßwerte vermissen. Besteht nicht die Möglichkeit, daß statt - oder zumindest neben - der Kreativität auch das Bedürfnis nach sozialer Anerkennung durch den freundlichen und allein mit dem jeweiligen Probanden beschäftigten Testleiter erfaßt wird? Cronbach (1968) bezeichnet die durch den Test gemessene Dimension als "responsiveness"; Guilford (1971) vermutet, daß das Verfahren wegen der fehlenden Zeitbegrenzung Persistenz oder Widerstand gegen Langeweile mißt. Da die Autoren das Problem alternativer Interpretationsmöglichkeiten offenbar nicht sehen, treffen sie auch keine Vorkehrungen zur Absicherung ihrer eigenen Interpretation.
Abgesehen von dem Einwand gegen die Forderung eines intelligenzunabhängigen Kreativitätstests und der Kritik wegen der fehlenden Berücksichtigung alternativer Interpretationsmöglichkeiten der Testergebnisse werden die dargestellten Argumente keineswegs als gewichtige Einwände gegen die Kreativitätstheorie von Wallach und Kogan betrachtet.
Ob das aus dieser Theorie entwickelte Verfahren den wichtigsten Testgütekriterien genügt, soll als nächstes untersucht werden.

2.32 Testgütekriterien

Bevor die Untersuchungsergebnisse zu den Testgütekriterien dargestellt werden, soll das Verfahren kurz beschrieben und einige Angaben zur Durchführung und Auswertung gemacht werden. Der Kreativitätstest von Wallach und Kogan besteht aus folgenden 5 Untertests:

1. "Instances" (4 Items): Die Kinder werden aufgefordert, alle möglichen Dinge mit einer bestimmten Eigenschaft - z.B. runde Dinge, Dinge, die Krach machen - aufzuzählen.

2. "Alternate Uses" (8 Items): Dieser von Guilford übernommene Aufgabentyp verlangt von den Kindern die Angabe möglichst vieler Verwendungsmöglichkeiten für alltägliche Gebrauchsgegenstände wie z.B. eine Zeitung, einen Knopf u.ä.

3. "Similarities" (1o Items): Die Kinder werden aufgefordert, zu sagen, in welcher Hinsicht sich 2 Dinge - z.B. Kartoffel und Mohrrübe, Eisenbahn und Traktor - ähneln. Auch auf diese Frage sollen - im Gegensatz zu der Aufgabe "Gemeinsamkeiten finden" im Hamburg-Wechsler-Intelligenztest - möglichst viele Antworten gegeben werden.

4. "Pattern Meanings" (8 Items): Der erste der beiden Untertests mit visuellen Aufgabenstimuli verlangt von den Kindern Auskunft darüber, was die 8 Zeichnungen einfacher geometrischer Figuren darstellen könnten. Durch ein Beispiel-Item werden sie dazu angeregt, möglichst viele und möglichst verschiedenartige Deutungen zu geben.

5. "Line Meanings" (9 Items): Dieselbe Aufgabenstellung wird für 9 Zeichnungen verschiedener Linien gegeben.

Wallach und Kogan untersuchten ihre Stichprobe von 151 Fünftkläßlern - 7o Jungen und 81 Mädchen - mit insgesamt 17 Verfahren; die 5 Kreativitäts-Untertests wurden an erster (Instances), fünfter (Pattern Meanings), sechster (Alternate Uses), neunter (Similarities) und zehnter Stelle (Line Meanings) gegeben. Den Kindern wurde gesagt, daß zwei Frauen (die beiden Testleiterinnen) verschiedene Spiele mit ihnen durchführen würden, um herauszufinden, ob ihnen diese Spiele gefallen.
Wie bereits erwähnt, wurden die von den Kindern mündlich gegebenen Antworten nach Anzahl und Einzigartigkeit ausgewertet. Wallach und Kogan betonen, daß die spielähnliche Testsituation nicht zu einer Überschreitung der durch die jweilige Aufgabenstellung gegebenen Grenzen geführt habe; bizarre oder inadäquate Antworten seien praktisch nicht aufgetreten. Dazu ist zu sagen, daß entweder die Auswertung bei Wallach und Kogan überaus großzügig war oder daß sich die Individualtestung qualitätssteigernd ausgewirkt hat, denn in der Untersuchung von Krause (1972), der eine Gruppentestung durchführte, sind Antworten aufgetreten, die zwei zuerst unabhängig und dann gemeinsam arbeitende Auswerter nicht in Zusammenhang mit den Aufgabenstimuli bringen konnten.
Cronbach (1971, zitiert nach Crockenberg 1972) scheint die erstgenannte Möglichkeit für zutreffend zu halten, denn er spricht von "failure of Wallach and Kogan to impose a quality criterion in

scoring..." (Crockenberg 1972, S. 36/37). Dieser Ansicht ist nicht unbedingt zuzustimmen; wenn Erwachsene Kreativitätstests von Kindern auswerten, ist wegen der Distanz zu den kindlichen Assoziationsreservoirs eine großzügige Auswertung sicher angebrachter als eine strenge.
Für den Gesamttest von Wallach und Kogan ergeben sich für die 39 nach Anzahl und Einzigartigkeit[1] ausgewerteten Items 78 Einzelscores, die zu je 5 Untertestscores für die Anzahl der Lösungen - im folgenden auch als "Produktivität" bezeichnet - und ihre Einzigartigkeit - in Anlehnung an Guilford und Torrance auch "Originalität" genannt - aufsummiert werden.
Die für die Testgütekriterien relevanten Untersuchungsergebnisse werden im folgenden dargestellt.

2.321 Objektivität

Da Wallach und Kogan ihr Verfahren für Forschungszwecke und in erster Linie für den eigenen Gebrauch entwickelten, befaßten sie sich kaum mit der Gewährleistung der verschiedenen Objektivitätsarten.
Während die Durchführungsobjektivität durch ihre Angaben zur Testdurchführung und die wörtliche Wiedergabe der Instruktion hinreichend gesichert erscheint, dürfte sich das Fehlen detaillierter Auswertungsanweisungen auf Replikationsversuche oder weiterführende Untersuchungen ungünstig auswirken. Wallach und Kogan ließen die 151 Tests ihrer Untersuchung durch 2 unabhängig arbeitende Mitarbeiter auswerten und geben an, 98-99 Prozent Übereinstimmung in den Auswertungen gefunden zu haben; Erläuterungen, wie dies festgestellt wurde, fehlen.
Kogan und Morgan (1969) überprüften ebenfalls die Auswerterobjektivität. Die Produkt-Moment-Korrelation zwischen den Produktivitätsscores zweier Auswerter für die Antworten von 1o9 Fünftkläßlern in den Alternate-Uses- und Pattern-Meanings-Aufgaben betrug .99; die Koeffizienten für die Originalitätswerte waren .82 für Alter-

1) Ein Punkt für "uniqueness" wurde von Wallach und Kogan für Antworten gegeben, die in ihrer Stichprobe nur ein einziges Mal aufgetreten sind, während Krause mit dem Hinweis auf seine größere Stichprobe auch für zweimal vorkommende Lösungen Originalitätspunkte vergibt.

nate Uses und .92 für Pattern Meanings. Nicholls (1971) erzielte noch höhere Werte: r=.99 für die Auswertungsobjektivität der Anzahl-, r=.95-.98 für diejenige der Originalitätswerte. Ward, Kogan und Pankove (1972) ermittelten nur Produktivitätsscores und stellten eine Auswertungsobjektivität von .94 für 4 Aufgaben mit je 4 Items fest.

Interpretationshilfen werden nicht gegeben; sie erscheinen den beiden Autoren vermutlich überflüssig, weil sie sowohl Anzahl als auch Originalität der Antworten als Maße des kreativen Denkens betrachten und eine relativ hohe Korrelation zwischen den beiden Kreativitätsoperationalisierungen erwarten. Die Frage, wie ein starker Unterschied zwischen den beiden Maßen zu interpretieren ist, wird nicht erörtert, obwohl Wallach und Kogan in ihrem Buch 32 Einzelfälle darstellen und es bei der Höhe der tatsächlich gefundenen Korrelationen zwischen Produktivität und Einzigartigkeit durchaus möglich ist, daß unter diesen Kindern einige waren, die hohe Anzahl- und geringe Originalitätswerte hatten.[1)]

2.322 Reliabilität

Während der Auswertungsobjektivität wenig Beachtung geschenkt wurde, ist die Interne-Konsistenz-Reliabilität von zentraler Bedeutung für die Konstruktvalidierung des Verfahrens von Wallach und Kogan. Bevor die dazu vorliegenden Ergebnisse dargestellt werden, soll noch kurz die wenig untersuchte Frage der Retest-Reliabilität besprochen werden.

Retest-Reliabilität

Da für den Kreativitätstest von Wallach und Kogan keine Parallelform vorliegt, kann die Retest-Reliabilität nicht innerhalb eines Zeitraumes von einigen Monaten oder gar Wochen festgestellt werden, da vermutlich gerade kreative Probanden kaum aufgabenorientierte Motivationen für die zweimalige Bearbeitung derselben Aufgaben entwickeln dürften. Bei Ergebnissen von Testwiederholungen nach längeren Zeitintervallen wird jedoch das Reliabilitätsmaß durch die eventuelle Instabilität des gemessenen Merkmals

1) Wallach und Kogan fanden für ihre 5 Untertests Korrelationen zwischen Anzahl und Originalität zwischen .08 und .71 mit einem Median von .64. Krause (1972) errechnete Koeffizienten zwischen .54 und .71 (Md=.62), außerdem korrelierte er den Gesamtscore für Anzahl mit dem Gesamtscore für Originalität und erhielt einen Koeffizienten von .72.

beeinflußt. Bei einem Kreativitätstest, der besonders empfindlich ist gegen Änderungen der Motivation - entweder durch zwischenzeitliche Erfahrungen oder durch veränderte Testbedingungen bei der zweiten Testung -, kann daher sicher nicht mit hohen Retest-Reliabilitäten gerechnet werden.

Kogan und Pankove (1971) wiederholten an 2 Gruppen von Fünftkläßlern - 72 und 29 Pbn - Subtests des Verfahrens von Wallach und Kogan nach einem Intervall von 5 Jahren. Die für Jungen und Mädchen getrennt ermittelten Ergebnisse werden in Tabelle 3 dargestellt.

Tabelle 3: Retest-Reliabilitäten (Produkt-Moment-Korrelationen) zwischen den Produktivitäts- und Originalitätswerten zweier Subtests von Wallach und Kogan für ein Zeitintervall von 5 Jahren

Schule A (16 Jungen, 13 Mädchen; Individualtestung)

		1.	2.	3.	4.	5.	6.
1. Produktivität	Alternate Uses (AU)	.o4)* / -.o2					
2. "	Pattern Meanings (PM)		.15 / .23				
3. "	AU + PM			.1o / .13			
4. Originalität	AU				.4o / .o7		
5. "	PM					.32 / .34	
6. "	AU + PM						.39 / .35

Schule B (38 Jungen, 34 Mädchen; Gruppentestung)

		1.	2.	3.	4.	5.	6.
1. Produktivität	Alternate Uses (AU)	.52ss / .o5					
2. "	Pattern Meanings (PM)		.39ss / .16				
3. "	AU + PM			.55ss / .o7			
4. Originalität	AU				.38ss / -.o9		
5. "	PM					.42ss / .23	
6. "	AU + PM						.49ss / .19

*) In der oberen Zeile stehen die Koeffizienten der Jungen, in der unteren die der Mädchen. Die Irrtumswahrscheinlichkeit von .o5 ist durch s, diejenige von .o1 durch ss gekennzeichnet.

Insgesamt sind die in Tabelle 3 dargestellten Koeffizienten sehr niedrig, signifikante Korrelationen ergaben sich nur für die 38 Jungen, mit denen Gruppentestungen durchgeführt worden waren. Da die als nächstes darzustellenden Koeffizienten der internen Konsistenz wesentlich höher liegen, ist anzunehmen, daß das durch die beiden Subtests gemessene Merkmal keine hohe zeitliche Stabilität besitzt. Die alternative Erklärung, daß die gleichen Aufgaben eines Kreativitätstests nicht ein zweites Mal durchgeführt werden können, erscheint bei einem Zeitintervall von 5 Jahren wenig plausibel.

Interne Konsistenz

Als erstes berechnen Wallach und Kogan für die 5 Produktivitäts- und die 5 Originalitätsscores jedes Subtests die Split-half-Reliabilität nach Spearman-Brown und erhalten Koeffizienten zwischen .51 und .93, der Median ist .875. Sie stellen fest, daß jedes ihrer 1o Kreativitätsmaße eine hohe innere Konsistenz besitzt.

Die zweite Reliabilitätsprüfung besteht in einer Trennschärfebestimmung für die Items der 5 Subtests. Für die insgesamt 39 Anzahl- und 39 Originalitätswerte wird die Korrelation mit dem Subtest-Anzahl- oder Subtest-Originalitätswert berechnet. Alle Koeffizienten liegen über .4o, 71 der 78 über .6o; somit hat auch die Überprüfung der Item-Trennschärfen gute Resultate erbracht.

Die zentrale Frage der Reliabilitätsuntersuchung ist jedoch diejenige nach den Interkorrelationen der 1o Kreativitätsmaße.

In der Untersuchung von Wallach und Kogan ergaben sich für die 45 Interkorrelationen der Produktivitäts- und Originalitätswerte der 5 Subtests 41 hochsignifikante und 2 signifikante Koeffizienten; die Werte liegen zwischen .o7 und .74, der Median beträgt .41. Nachdem die beiden Autoren Vorkehrungen getroffen haben, damit die Zusammenhänge nicht wegen der hohen Korrelationen zwischen Anzahl und Einzigartigkeit derselben Aufgabe überschätzt werden,[1] kann die Höhe der Koeffizienten als zufriedenstellend bezeichnet werden. Bemerkenswert ist, daß die Subtests mit verbalen und diejenigen mit visuellen Aufgabenstimuli untereinander ebenfalls überwiegend hochsignifikant und z.T. substantiell korrelieren.

---

1) Bei den für die Interkorrelationsberechnungen verwendeten Anzahl-Scores wurden die nur einmal aufgetretenen Antworten nicht mitgezählt.

Die Zusammenhänge zwischen den um die Zahl der einzigartigen Antworten verminderten Anzahlwerten und den Originalitätsscores sind recht unterschiedlich: für Instances beträgt die Korrelation .o8, für Alternate Uses .67, für Similarities .71, für Pattern Meanings .29 und für Line Meanings .64. Hier stellt sich die Frage, ob diese Ergebnisse mit dem Modell in Einklang gebracht werden können.
Krause (1972), der den Test von Wallach und Kogan mit 2o2 Viertkläßlern durchführte, erhält vergleichbare Ergebnisse (.19 ≤ r ≤ .71, Md.=.38). Er gibt für seine Koeffizienten einen Gesamtdurchschnitt von .41o2 an, für diejenigen von Wallach und Kogan beträgt der entsprechende Wert .4o6.[1] Wenn man berücksichtigt, daß Krause seine Produktivitätswerte nicht um die Anzahl der einzigartigen Antworten verringerte,[2] sind die Interkorrelationen seiner Untersuchung insgesamt etwas niedriger, der Unterschied ist aber so gering, daß der nach der Theorie der Testautoren zu vermutende Grund - die klassenweise Testdurchführung mit einem den Kindern nicht bekannten Testleiter - sich nur sehr schwach ausgewirkt haben kann.
Die interne Konsistenz des Kreativitätstests von Wallach und Kogan wurde in einer ganzen Reihe von Untersuchungen mit Probanden derselben Altersstufe,[3] jüngeren[4] und älteren Kindern[5] und Erwachsenen[6] überprüft (meist aufgrund von Gruppentestungen); in nur einer Untersuchung wurde das Verfahren unverändert[7], in allen anderen wurde es gekürzt und/oder mit kleineren Modifikationen übernommen, aber die Ergebnisse sprachen fast immer für eine innere Konsistenz der Subtests, die derjenigen, die die Testautoren selbst ermittelt haben, zumindest gleichkommt[8]. In drei Untersuchungen wurde festgestellt, daß die Originalitätsscores geringere Reliabilitäten besitzen als die Anzahlwerte[9].

---

1) S. Krause 1972 , S. 1o3

2) Vgl. Anmerkung 1), S. 91

3) Grote, Hajek u.a. 1969, Hargreaves and Bolton 1972, Kogan and Pankove 1971, Leith 1972, Nicholls 1971, Ward 1971, Ward, Kogan and Pankove 1972

4) Rotter, Langland and Berger 1971, Williams and Fleming 1969

5) Kogan and Pankove 1971, Murphy 1973

6) Cropley 1968, Cropley and Maslany 1969, Richards 1974

7) Cropley 1968

8) Drei Untersuchungen erbrachten z.T. abweichende Resultate: Ward (1968) für 87 Vier- bis Sechsjährige, Massad (1969) für 39 Sechstkläßler mit niedrigerem sozioökonomischen Status und Kogan und Morgan (1969) für 25 Jungen, deren Flüssigkeitsscores zweier Subtests unter spielähnlichen Testbedingungen nicht korrelierten.

9) Murphy 1973 , Richards 1974 , Wagner 1972

Die meisten der hier berücksichtigten Untersuchungsresultate sprechen jedoch für die innere Konsistenz sowohl der Item- als auch der Subteststscores beider Kreativitätsmaße und rechtfertigen damit die Bildung eines "Kreativitäts-Gesamtscores" aus den 1o Produktivitäts- und Originalitätswerten.

2.323 Validität

Das Konzept der Inhaltsvalidität ist für Wallach und Kogan nicht relevant, da sie ihre Testaufgaben nicht als Stichprobe kreativer Verhaltensweisen betrachten und die Gültigkeit ihres Verfahrens im Rahmen einer Konstruktvalidierung überprüfen. Da die Testautoren selbst keine kriterienbezogenen Validitätsuntersuchungen durchführten, sollen zuerst die Untersuchungen zur Konstruktvalidität dargestellt werden.

2.3231 Konstruktvalidität

Die bereits dargestellten Ergebnisse zur internen Konsistenz des Verfahrens sind auch für den Nachweis der Konstruktvalidität von großer Bedeutung, da Wallach und Kogan von einem Kreativitätstest fordern, daß er eine ähnlich einheitliche Persönlichkeitsdimension erfaßt wie ein Intelligenztest.

Nachdem die Resultate zur internen Konsistenz des Tests gut ausfielen, stellte sich als nächstes die ebenso wichtige Frage nach der Unabhängigkeit der Testergebnisse von der allgemeinen Intelligenz. Wallach und Kogan stellten ihren 1o Kreativitätsmaßen - den Produktivitäts- und Originalitätsscores der 5 Untertests - 1o Intelligenztestwerte gegenüber,[1] die einen noch engeren Zusammenhang untereinander zeigen als die 1o Kreativitätsmaße. Die Korrelationen der Kreativitäts- mit den Intelligenztestergebnissen liegen zwischen .13 und .23 (Md=.1o), sind überwiegend insignifikant (79 von 1oo Koeffizienten) und bestätigten somit in vollem Umfang die Erwartungen der Testautoren.

Der Kreativitäts-Intelligenz-Zusammenhang, untersucht mit dem - z.T. gekürzten und/oder modifizierten - Verfahren von Wallach und Kogan und einer ganzen Reihe von Intelligenztests, wurde in einer größeren Zahl weiterer Arbeiten mit Stichproben verschiedener Altersstufen

---

1) Es handelt sich um 3 Subtests der Wechsler Intelligence Scale for Children, den verbalen und numerischen Gesamtwert des School and College Ability Tests und die 5 Subtests der Sequential Tests of Educational Progress. Cronbach (1968) wendet gegen diese Zusammenstellung zu Recht ein, daß es sich überwiegend um Schulleistungstests handele; auch einer der Wechsler-Subtests (Wortschatz) erfaßt eine stark schulleistungsrelevante Fähigkeit.

überprüft.[1] Bestätigungen der Ergebnisse von Wallach und Kogan erbrachten die Untersuchungen von Cropley (1968), Cropley und Maslany (1969), Kogan und Morgan (1969), Leith (1972), Massad (1969), McKinney und Forman (1977), Murphy (1973), Nicholls (1971), Richards (1974), Rotter, Langland und Berger (1971), Wagner (1972), Ward (1971), Ward, Kogan und Pankove (1972) und Williams und Fleming (1969). Auch faktorenanalytische Untersuchungen der Daten von Wallach und Kogan durch Fee (1968) und Ward (1967, zitiert nach Kogan 1971) erbrachten klare Trennungen der Kreativitäts- von den Intelligenz-Subtests. Teilweise widersprüchliche faktorenanalytische Ergebnisse von Cropley (1968) und Cropley und Maslany (1969) werden von Kogan (1971) dadurch erklärt, daß die Faktorenmatrizen nicht rotiert worden waren; nach Rotation der Faktorenmatrix von Cropley und Maslany erhält Kogan eine klare Trennung der Kreativitäts- von den Intelligenzvariablen.

Die Arbeit von Kogan und Pankove (1971) bestätigte für die Probanden der fünften Klassen die Unabhängigkeit der Kreativitäts- von den Intelligenztestergebnissen; bei den Zehntkläßlern ergaben sich für Jungen und Mädchen und Gruppen- versus Individualtestung unterschiedliche Ergebnisse, die die Theorie Wallachs und Kogans über die Auswirkung der Testsituation nur teilweise bestätigen.

Ward (1968) erhielt für 34 sieben- und achtjährige Schüler insignifikante Korrelationen zwischen Kreativitäts- und Intelligenzmaßen, aber die 41 Jungen seiner 4- bis 6jährigen Probanden hatten verbale Kreativitätsscores, die signifikant negativ mit dem Peabody Picture Vocabulary Test korrelierten, während die beiden Scores für Pattern Meanings positive Beziehungen zum PPVT hatten.

In den beiden mit deutschen Schülern vierter Klassen durchgeführten Untersuchungen wurden die Ergebnisse von Wallach und Kogan nicht bestätigt: bei Krause (1972) korreliert der Gesamtwert aus allen 10 Maßen des Verfahrens von Wallach und Kogan mit o.44 mit dem Leistungs-Prüf-System von Horn; Grote, Hajek u.a. (1969) erhielten für den Zusammenhang einzelner Aufgaben dieses Kreativitätstests mit dem Frankfurter Analogietest eine durchschnittliche Rangkorrelation von .33.

---

1) Ungefähre Angaben zum Alter der Probanden werden bei der Darstellung der Untersuchungen zur internen Konsistenz gegeben, s. S. 92 . Die Stichprobe der dort nicht berücksichtigten Arbeit von Kogan und Morgan (1969) bestand aus Fünftkläßlern, diejenige der Untersuchung von Wagner (1972) aus High School Seniors.

Auch die mit 117 englischen Elementarschülern durchgeführte Untersuchung von Hargreaves und Bolton (1972) erbrachte hochsignifikante und substantielle Korrelationskoeffizienten für die Beziehung zwischen den Anzahl-Scores zweier Subtests von Wallach und Kogan und 2 Intelligenzmaßen ($.36 \leq r \leq .53$).
Es liegt nahe, die abweichenden Untersuchungsergebnisse nach der Theorie von Wallach und Kogan durch die allen drei Untersuchungen gemeinsame Gruppentestung zu erklären. Dem widerspricht jedoch die Tatsache, daß in mehreren Untersuchungen, die die Unabhängigkeit der Kreativitäts- von den Intelligenztestergebnissen bestätigten, auch Gruppentests, z.T. sogar mit Zeitbegrenzung, durchgeführt wurden (Leith 1972, Richards 1974, Ward 1971 mit, Cropley 1968, Cropley und Maslany 1969 ohne Zeitbegrenzung). Auch in der Arbeit von Kogan und Morgan (1969) wurden trotz Gruppentestung mit Zeitbegrenzung kaum signifikante Beziehungen zwischen Kreativitäts- und Intelligenzmaßen gefunden (von 48 Koeffizienten war nur einer signifikant).
Kogan und Morgan versuchen, dieses der Theorie von Wallach und Kogan widersprechende Ergebnis durch Besonderheiten ihrer Stichprobe zu erklären, die - im Gegensatz zu den bisher fast ausschließlich untersuchten Kindern aus mittlerer und oberer Mittelschicht - aus Arbeiterfamilien und Familien der unteren Mittelschicht kam.
Es ist allerdings auch möglich, daß die bisher verwendeten Stichproben aus mittlerer und oberer Mittelschicht aus überdurchschnittlich intelligenten Kindern bestanden, für die der Zusammenhang zwischen Kreativität und Intelligenz weniger eng zu sein scheint als für weniger intelligente Kinder (s. dazu S. 73-74). Krause (1972), der die Kreativitäts-Intelligenz-Korrelationen getrennt für die 4 Quartile der Intelligenz-Verteilung seiner Probanden berechnete, fand für das zweite Quartil die engsten Zusammenhänge, und Hargreaves und Bolton (1972) stellten für ihre untere und mittlere Intelligenzgruppe eine hochsignifikante und relativ enge Beziehung zwischen Kreativitäts- und Intelligenzmaßen fest, während die Korrelation für die Gruppe der Hochintelligenten (IQ-Mittelwert =122).o9 betrug. Im Gegensatz zu Krause führen sie dieses Ergebnis auf die bei höherer Intelligenz zu erwartende Unabhängigkeit von Kreativität und Intelligenz und nicht auf die eingeschränkte Streuung der Intelligenztestwerte zurück.
Diese dürfte sich bei den dargestellten Widersprüchen der Untersuchungsergebnisse allerdings zusätzlich ausgewirkt haben. Krause

weist ausdrücklich auf die große Intelligenzstreuung seiner Stichprobe hin, ebenso Hargreaves und Bolton (1972); die Stichprobe von Grote, Hajek u.a. (1969) stammt aus einer 4.Grundschulklasse, in der noch keine Aufteilung der Kinder auf verschiedene Schultypen vorgenommen wurde, so daß auch hier von einer weiten Streuung der Intelligenzwerte ausgegangen werden kann.

Wallach und Kogan machen keine Angaben über die Verteilung der Intelligenztestergebnisse in ihrer Untersuchung; aufgrund der mitgeteilten Berufe der Väter ihrer Probanden kann jedoch angenommen werden, daß ihre Stichprobe überdurchschnittlich intelligent war[1] und eine eingeschränkte Intelligenzstreuung besaß.
Somit kann zusammenfassend nur gesagt werden, daß die von Wallach und Kogan gefundene Unabhängigkeit der Kreativitäts- von den Intelligenztestergebnissen nicht durch alle anderen hier berücksichtigten Untersuchungen bestätigt wurde und daß es möglich ist, daß das Ergebnis der Testautoren nur für Stichproben mit höherer Intelligenz gültig ist. Ihre Hypothese, daß die Unabhängigkeit der Kreativitätsmaße von der Testintelligenz nur nachgewiesen werden kann, wenn erstere in einer spielähnlichen Testsituation in Einzeluntersuchung und ohne Zeitbegrenzung erhoben werden, wurde durch mehrere Untersuchungen widerlegt.
Welche Bedeutung haben diese Ergebnisse für die Konstruktvalidität des Verfahrens von Wallach und Kogan? Die zentrale Forderung der beiden Autoren an einen Kreativitätstest war, daß das Verfahren eine einheitliche und von Intelligenztestergebnissen unabhängige Persönlichkeitsdimension erfassen soll; diese Forderung sollte als erfüllt betrachtet werden, wenn die Subtests des Verfahrens zur Messung der Kreativität untereinander hoch und mit Intelligenz-Subtests nicht oder nur schwach korrelieren. Ihr Hauptargument gegen die vorliegenden Kreativitätstests war, daß die einzelnen Aufgaben untereinander etwa ebenso hoch wie mit Intelligenztestergebnissen zusammenhängen, so daß die Möglichkeit nicht auszuschließen ist, daß

---

1) Es handelt sich fast ausschließlich um Kinder von Geschäftsleuten, Angestellten und Akademikern und somit um Probandengruppen, die in den meisten Gruppentests zur Intelligenzmessung besonders gut abschneiden (s. dazu z.B. Eells 1951, zitiert nach Oevermann 1971[6]).

die Aufgaben nur aufgrund ihrer gemeinsamen Beziehung zur Testintelligenz korrelieren.

Diese Möglichkeit besteht bei den 1o Kreativitätsmaßen von Wallach und Kogan jedoch nicht. Auch in den 3 Untersuchungen, in denen die Kreativitäts-Subtests signifikant mit Intelligenztestwerten korrelierten, übertrifft die interne Konsistenz der ersteren bei weitem ihre Korrelationen mit letzteren.[1)] Somit erfüllt das Verfahren von Wallach und Kogan die wichtigste Forderung der Testautoren, auch wenn von einer völligen Unabhängigkeit seiner Ergebnisse von der Testintelligenz nicht ausgegangen werden kann.
(Auf die Problematik eines von der Intelligenz völlig unabhängigen Kreativitätskonzepts wurde bereits hingewiesen, vgl. S. 85.)
Als nächstes soll überprüft werden, ob die aus der assoziativen Kreativitätstheorie Mednicks und Wallachs und Kogans abzuleitenden Hypothesen durch die Untersuchungsergebnisse bestätigt wurden oder nicht.
Die nach Mednicks Modell sehr enge Beziehung zwischen Anzahl und Einzigartigkeit der Assoziationen wurde bestätigt. Wallach und Kogan (1965) fanden wegen der Reduktion der Produktivitätsscores um die Zahl der einzigartigen Antworten die niedrigsten Korrelationen ($.08 \leq r \leq .71$, Md=.64); die Mediane der Koeffizienten anderer Untersuchungen liegen höher (Grote, Hajek u.a. 1969: Md=.88; Krause 1972: Md=.72; Murphy 1973: Md=.71). Ward (1968) fand für 3 Subtests von Wallach und Kogan, mit denen er 34 7- bis 8jährige Jungen untersuchte, so hohe Korrelationen zwischen Anzahl- und Originalitätswerten (.92, .94 und .98), daß er in einer späteren Untersuchung (Ward 1971) auf die Originalitätsauswertung verzichtete.

Aus dem Modell von Mednick können einige weitere Hypothesen abgeleitet werden:

1. Die Antwort-Rate (Zahl der Antworten pro Zeiteinheit) nimmt im Verlauf der Aufgabenbearbeitung ab.

---

1) Bei Grote, Hajek u.a. (1969) korrelieren die 14 Anzahl- und Originalitätswerte der 7 Items von Wallach und Kogan zwischen .o8 und .97 untereinander, der Median ist .54; mit dem IQ des Frankfurter Analogietests korrelieren sie zwischen .1o und .58, Md=.295. Hargreaves und Bolton teilen nur die Korrelationen der Flüssigkeitsscores zweier Kreativitätssubtests untereinander (r=.7o) und mit 2 Intelligenztestwerten mit ($.36 \leq r \leq .53$). Krause (1972) macht zwar keine Angaben über die Beziehungen zwischen Kreativitäts- und Intelligenz-Subtests, aber seine Gesamtwerte für Produktivität und Originalität korrelieren mit den Subtests des Leistungs-Prüf-Systems von Horn niedriger: $.11 \leq r \leq .43$, Md=.2o) als die 1o Kreativitätsmaße untereinander ($.19 \leq r \leq .71$, Md=.38).

2. Ungewöhnliche bzw. einzigartige Antworten nehmen im Verlauf der Aufgabenbearbeitung zu.

3. Kreative Individuen geben insgesamt eine größere Zahl von Antworten und eine größere Zahl einzigartiger Antworten als weniger kreative.

4. Kreative Individuen haben zu Beginn der Aufgabenbeantwortung eine niedrigere Antwort-Rate als weniger kreative.

5. Kreative benötigen mehr Zeit zur Aufgabenbearbeitung als Unkreative.

Die ersten beiden Hypothesen wurden von Ward (1969a) für 7- bis 8jährige Jungen mit Hilfe dreier Subtests von Wallach und Kogan bestätigt; Ward zitiert außerdem Untersuchungen mit College-Studenten, die ebenfalls die Richtigkeit der beiden Annahmen belegen. Die 2. Hypothese wurde zusätzlich durch Ergebnisse der Untersuchung von Grote, Hajek u.a. (1969) bestätigt.
Die 3. Hypothese kann nur mit Hilfe eines Außenkriteriums, das eine Identifikation kreativer und weniger kreativer Individuen ermöglicht, nicht jedoch allein durch eine Analyse der Testergebnisse überprüft werden. Einige später zu besprechende Untersuchungen zur Übereinstimmungs- und Vorhersagevalidität bestätigen die Validität des Verfahrens von Wallach und Kogan bis zu einem gewissen Grad und stützen damit auch die 3. Hypothese. (Daß der durch die Hypothese postulierte enge Zusammenhang zwischen Zahl und Einzigartigkeit der Antworten mehrfach belegt ist, wurde bereits besprochen, s.o.)
Die vierte Hypothese kann z.T. durch Vergleich der Antwortzahlen in verschiedenen Zeitintervallen während der Aufgabenbearbeitung überprüft werden; im ersten Drittel, Viertel o.ä. der Bearbeitungszeit müßten die Probanden, die überdurchschnittlich viele oder überdurchschnittlich viele einzigartige Lösungen anbieten, weniger Antworten geben als diejenigen, die insgesamt weniger produzieren. (Außer dieser Überprüfung wäre dann noch der Nachweis zu erbringen, daß diejenigen, die viele und viele originelle Lösungen liefern, auch wirklich kreativ sind!)
Die Ergebnisse der Untersuchung von Ward (1969a) bestätigen diese Hypothese nicht: diejenigen Jungen, die insgesamt hohe Anzahl- und Originalitätswerte hatten, gaben ihre Antworten während der gesamten Bearbeitungsdauer mit höherer Rate als die weniger produktiven Jungen. Damit ist die auch von Wallach und Kogan vermutete inverse Beziehung

zwischen der Anzahl der Assoziationen und ihre Produktionsrate widerlegt (vgl. S. 79 und 81).
Die letzte Hypothese erfordert ebenfalls die Verwendung eines zweiten Verfahrens zur Feststellung der Kreativität der Probanden. Es scheinen bisher kaum Untersuchungen mit dem Verfahren von Wallach und Kogan vorzuliegen, in denen lediglich die Testzeit variiert wurde, so daß noch keine gesicherten Aussagen über die Richtigkeit dieser Hypothese, deren Bedeutung von Wallach und Kogan selbst betont wurde (vgl.S.81), möglich sind. Die Ergebnisse mehrerer Untersuchungen, in denen der Test im Gruppenrahmen mit und ohne Zeitbegrenzung durchgeführt wurde, waren z.T. widersprüchlich und erlauben daher noch keine sicheren Schlußfolgerungen. (Sie werden unter Punkt 2.331 dargestellt, s. vor allem S. 125-126.)

Insgesamt betrachtet stützen die Untersuchungsergebnisse von Wallach und Kogan und einer Reihe weiterer Forscher die assoziative Kreativitätstheorie von Mednick in mehreren wichtigen Punkten. Kleinere Abweichungen wie die positive Beziehung zwischen Antwortzahl und Produktionsrate müßten noch einmal überprüft werden und würden im Fall ihrer Bestätigung nur eine kleinere Modifikation des Modells von Mednick erfordern.
Neben dieser ausschließlich auf Mednicks Theorie bezogenen Konstruktvalidierung im engeren Sinn führten Wallach und Kogan weitere Untersuchungen zu einer größeren Anzahl von Verhaltensbereichen durch, die als Konstruktvalidierung ihres Kreativitätstests im weiteren Sinn[1] betrachtet werden können.
Die Untersuchungen zu Angst und Abwehrtendenz sind durch Hypothesen über die optimale Testsituation relativ eng mit der assoziationstheoretischen Kreativitätskonzeption verbunden (s. dazu S. 83-84) und sollen daher zuerst besprochen werden.
Wallach und Kogan erfaßten allgemeine Angst, Test- bzw. Schulangst[2] und Abwehrtendenz durch Adaptationen der Fragebogen von Sarason,

---

1) Diese Konstruktvalidierung i.w.S. basiert allerdings auf theoretisch und empirisch sorgfältig begründeten Hypothesen und ist somit nicht mit den Untersuchungen zu vergleichen, die Torrance (1966, 1967a)unter der Überschrift "Konstruktvalidierung" referiert (vgl.S. 5o-55).

2) Da sich die Fragen auf schulische Situationen beziehen, erscheint die Bezeichnung "Schulangst" treffender als die von Wallach und Kogan verwendete Bezeichnung "test anxiety".

Davidson et al. (196o, zitiert nach Wallach und Kogan 1965).
Da die Mittelwerte der Jungen und Mädchen in allgemeiner Angst und Abwehrtendenz signifikant verschieden waren,[1] wurden die Berechnungen - wie auch alle später zu besprechenden Untersuchungen - für die beiden Geschlechter getrennt durchgeführt.

Unter der Fragestellung, wie sich Intelligenz und Kreativität auf die Persönlichkeitsdimensionen allgemeine Angst, Schulangst und Abwehrtendenz auswirken, bildeten Wallach und Kogan aufgrund von Medianhalbierungen nach den Intelligenz- und Kreativitätstestergebnissen 4 Gruppen von Jungen und Mädchen:
a) Intelligente und kreative,
b) intelligente, aber wenig kreative,
c) wenig intelligente, aber kreative und
d) wenig intelligente und wenig kreative
Jungen bzw. Mädchen.

Unter veränderter Fragestellung wurden Schulangst und Abwehrtendenz als unabhängige Variablen betrachtet und ihre Effekte auf Intelligenz und Kreativität untersucht, wobei Medianhalbierungen der beiden Stichproben aufgrund der Schulangst- und Angstabwehrwerte vorgenommen wurden.
Mit den Daten wurden Korrelations- und eine größere Zahl von Varianzanalysen gerechnet.
Nur folgende Ergebnisse haben für Jungen und Mädchen Gültigkeit:
Zwischen allgemeiner Angst und Schulangst bestand eine substantielle Beziehung um .6o. Der Zusammenhang zwischen allgemeiner Angst und Abwehrtendenz war negativ, aber nicht sehr hoch (-.24 für Jungen, -.32 für Mädchen).
Kinder mit hohen Werten für Angst und Angstabwehr hatten die schlechtesten Intelligenztestergebnisse.
Die niedrigsten Werte im Kreativitätstest hatten diejenigen Kinder, bei denen ein geringes Ausmaß an Schulangst mit einer starken Ab-

---

1) Jungen hatten niedrigere Angst- und höhere Angstabwehr-Werte als Mädchen, was aufgrund der geschlechtsrollenspezifischen Erziehung zu erwarten war. Sarason, Davidson et al. (196o, zitiert nach Wallach und Kogan 1965) weisen außerdem darauf hin, daß der Inhalt der Fragebogen-Items für Mädchen eine geringere Bedeutung habe als für Jungen und daher bei ersteren eine geringere Angstabwehr, verbunden mit höheren Angstwerten, hervorrufe.

wehrtendenz verbunden war; allerdings waren die Effekte weder für Schulangst noch für Abwehrtendenz signifikant. Nach Wallach und Kogan handelte es sich bei dieser Gruppe um die defensivsten Kinder - um diejenigen nämlich, die ihre Schulangst nicht zugeben. Daß diese Kinder die schlechtesten Kreativitätstestergebnisse haben, werten sie als Bestätigung ihrer Hypothese, daß zwischen Kreativität und Abwehrtendenz eine inverse Beziehung besteht. Diese Interpretation erscheint angesichts des für Jungen und Mädchen insignifikanten Effekts der Abwehrtendenz fragwürdig.

Alle übrigen Ergebnisse fielen nach Wallach und Kogan für Jungen und Mädchen unterschiedlich aus und werden daher getrennt dargestellt.
Für Jungen ergaben sich unter der Fragestellung, welche Auswirkungen Intelligenz und Kreativität auf die untersuchten Persönlichkeitsvariablen haben, folgende Resultate:
Intelligente, aber wenig kreative Jungen hatten die niedrigsten Angst- und Schulangstwerte, wenig intelligente und wenig kreative Jungen hohe Angst- und Schulangstwerte, und die beiden Gruppen kreativer Jungen zeigten ein mittleres Niveau an zugegebener allgemeiner Angst und Schulangst. Wallach und Kogan sind der Ansicht, daß die intelligenten und wenig kreativen Jungen den schulischen Anforderungen am besten entsprechen und aus diesem Grund relativ angstfrei sind.
Im Gegensatz zur Intelligenz war der Effekt der Kreativität nicht signifikant.
Auch für die Abwehrtendenz wurde ein signifikanter Intelligenzeffekt gefunden: Intelligente Jungen hatten niedrige Abwehr-Werte. Die Gruppe der intelligenten und kreativen Jungen hatte die niedrigsten Werte für Abwehrtendenz, aber der Kreativitätseffekt war wieder nicht signifikant. Somit muß die ursprüngliche Untersuchungshypothese dahingehend modifiziert werden, daß nur bei intelligenten Jungen Kreativität mit geringer Angstabwehr verbunden ist.
Bei den Mädchen zeigten sich weder für Intelligenz noch für Kreativität Zusammenhänge mit den Angst- und Angstabwehr-Werten.
Die Untersuchung der Auswirkungen von Schulangst und Abwehrtendenz auf Intelligenz- und Kreativitätstestergebnisse erbrachte folgende Resultate:
Jungen mit hohen Schulangst-Werten hatten niedrige Intelligenztestscores. Die besten Ergebnisse erzielten die wenig ängstlichen und

wenig defensiven Jungen,die schlechtesten Intelligenztestwerte hatten die Jungen mit hohen Werten für beide Persönlichkeitsmerkmale; allerdings war nur der Effekt der Schulangst signifikant.
Die Abwehrtendenz hatte dagegen eine stärkere (fast signifikante) Auswirkung auf die Kreativitätstestergebnisse als die Schulangst.
Sowohl bei den wenig als auch bei den stark defensiven Jungen waren höhere Schulangstwerte mit besseren Kreativitätstestergebnissen verbunden, der Schulangst-Effekt war allerdings nicht signifikant.
Bei den Mädchen wurden weder für Intelligenz noch für Kreativität signifikante Schulangst- und Angstabwehr-Effekte gefunden. Starke Schulangst und hohe Abwehr-Werte führten zu disproportional schlechten Intelligenztestergebnissen (die Interaktion war fast signifikant), während die Kreativitätstestscores dieser Gruppe höher lagen als diejenigen der 3 anderen Gruppen.
Insgesamt stimmten die Ergebnisse der Jungen wesentlich besser mit den Untersuchungshypothesen überein als diejenigen der Mädchen. So wurde für erstere die erwartete negative Korrelation zwischen Schulangst und Intelligenz gefunden. Die vermutete inverse Beziehung zwischen Abwehrtendenz und Kreativität wurde allerdings nur für die intelligenten Jungen bestätigt.
Wallach und Kogan führten zusätzlich zu den besprochenen noch eine ganze Reihe weiterer Untersuchungen - zum Verhalten der Kinder im Unterricht, zu Variablen des kognitiven Stils u.a. - durch.
Für die bereits beschriebenen 4 Gruppen unterschiedlich intelligenter und kreativer Kinder wurden Hypothesen über die Resultate dieser Untersuchungen aufgestellt, die auf bereits vorliegenden Theorien und/oder empirischen Ergebnissen basieren.
Die nach Geschlechtern getrennte varianzanalytische Verarbeitung der Daten erbrachte für die Intelligenz eine wesentlich größere Zahl von signifikanten Haupteffekten als für die Kreativität. Im Zusammenhang mit der Fragestellung der vorliegenden Arbeit erscheinen vor allem die Untersuchungsergebnisse zum schulischen Leistungs- und Sozialverhalten - erfaßt durch Verhaltensratings zweier unabhängiger Beobachter - interessant:
Bei den Jungen ergaben sich nur für die Intelligenz signifikante Effekte. Die weniger intelligenten Jungen konnten sich weniger gut konzentrieren, zeigten weniger Interesse an der Schule und waren sehr intrapunitiv.

Bei den Mädchen zeigten sich dagegen Intelligenz- und Kreativitätseffekte und Interaktionen zwischen beiden Variablen.
Die intelligenten und kreativen Mädchen zeichneten sich durch ein hohes Maß an Ich-Stärke, Selbstvertrauen und aktiver Gestaltung ihrer Sozialbeziehungen aus. Durch ihr ungehemmtes Verhalten im Unterricht wirkten sie manchmal störend; auf der Verhaltens-Dimension "attention-seeking" erreichten sie die höchsten Werte.
Für die intelligenten, aber wenig kreativen Mädchen bot die Schule wenig Probleme; Wallach und Kogan sehen in den Ergebnissen dieser Gruppe allerdings Anzeichen für eine gewisse Zurückhaltung im Einsatz ihrer intellektuellen Fähigkeiten und in ihrem Sozialverhalten.
Die wenig intelligenten und wenig kreativen Mädchen kompensierten ihre schlechten Schulleistungen durch befriedigende Beziehungen zu den Klassenkameraden und boten insgesamt weniger Verhaltensprobleme als die wenig intelligenten, aber kreativen Mädchen.
Diese letztgenannte Gruppe hat nach den Ergebnissen von Wallach und Kogan die größten Schwierigkeiten, sie schneiden im Hinblick auf Selbstvertrauen, Ich-Stärke etc., in schulischen Variablen wie Interesse für und Konzentration auf den Unterrichtsstoff, Einschätzung ihrer eigenen Leistungen u.ä. und auf dem Gebiet der Sozialbeziehungen am schlechtesten ab.

Insgesamt zeigen die Intelligenztestergebnisse auch für die Verhaltensbeurteilungen der Mädchen die überwiegende Zahl signifikanter Haupteffekte: 6 Intelligenz-Effekten stehen ein signifikanter Haupteffekt der Kreativität, 2 signifikante und 2 fast signifikante Wechselwirkungen gegenüber. Cronbach (1968), der wegen einiger methodischer Einwände die Daten von Wallach und Kogan reanalysierte,[1] kommt aufgrund seiner Ergebnisse zu folgendem Schluß: "My final impression is that the F variable[2] has disappointingly limited psychological significance. It can scarcely be considered a measure of ability or creativity; there is no evidence that high F children produce responses of superior quality in any situation." (Cronbach 1968, S. 509).

Einige weitere Untersuchungen, die Aussagen über die Konstruktvalidität des Verfahrens von Wallach und Kogan ermöglichen, sollen noch kurz besprochen werden.

---

1) Cronbach bezweifelt vor allem die Notwendigkeit einer nach Geschlechtern getrennten Datenverarbeitung. Seine mit den Ergebnissen der Gesamtstichprobe und aufgrund multipler Regressionsanalysen durchgeführten Berechnungen kommen in einigen Punkten zu anderen Resultaten als Wallach und Kogan.

2) Cronbachs "neutrale Bezeichnung" für die Ergebnisse des Kreativitätstests von Wallach und Kogan.

Die Arbeit von Kogan und Morgan (1969), in der z.T. dieselben Fragen untersucht werden wie in derjenigen von Wallach und Kogan, ist vor allem deshalb interessant, weil sie eine Stichprobe (1o4 Fünftkläßler, 52 Jungen und 52 Mädchen) mit anderem sozioökonomischem Hintergrund verwendet; während die von Wallach und Kogan untersuchten Kinder der oberen Mittelschicht entstammten, handelt es sich bei der Stichprobe von Kogan und Morgan überwiegend um Arbeiterkinder der unteren Mittelschicht. Außerdem vermeiden Kogan und Morgan eine Schwäche der Untersuchung von Wallach und Kogan: Der Versuch, die Überlegenheit eines spielähnlichen Kontextes bei der Durchführung zweier Subtests des Verfahrens von Wallach und Kogan ("Alternate Uses" und "Pattern Meanings") nachzuweisen, wurde mit Hilfe einer Experimental- und einer Kontrollgruppe realisiert; letztere wurde in einer testähnlichen Situation mit betonter Zeitbegrenzung und unter stressinduzierenden Hinweisen und Ermahnungen untersucht, erstere in entspannter, schul**unähnlicher Atmosphäre mit unmerklicher Zeitbegrenzung, aber in** Gruppen und nicht wie bei Wallach und Kogan in Einzeluntersuchungen.

Im Gegensatz zur Untersuchungshypothese erreichte die Experimentalgruppe keine höheren Kreativitätstestscores als die Kontrollgruppe; die Produktivitäts- und Originalitätswerte des Subtests "Alternate Uses" fielen sogar für die Kontrollgruppe signifikant besser aus. Die nach der Theorie Wallachs und Kogans zu erwartenden Interaktionen zwischen Schulangst und Abwehrtendenz zeigten sich vor allem in der testähnlichen Untersuchungssituation. Die Effekte der Abwehrtendenz waren stärker als diejenigen der Schulangst, sie wirkten sich allerdings nur bei Jungen in der erwarteten Weise - also negativ auf die kreative Produktion - aus. In Übereinstimmung mit den Untersuchungshypothesen schnitten die wenig schulängstlichen Kinder im testähnlichen Kontext besser ab, im Gegensatz zu den Erwartungen führte die spielähnliche Testsituation nicht zu dem umgekehrten Ergebnis einer Überlegenheit der schulängstlichen Kinder. Im Hinblick auf die Intelligenz-Kreativitätsbeziehung war erwartet worden, daß die Maße der beiden Variablen nur unter testähnlichen, nicht jedoch unter spielähnlichen Untersuchungsbedingungen korrelieren. Diese Hypothese wurde nicht bestätigt (s. dazu S. 95). Allerdings führte die Verwendung der Moderatorvariablen Schulangst für die Jungen zu Ergebnissen, die mit der Untersuchungshypothese übereinstimmen, d.h. in der test-

ähnlichen Untersuchungssituation korrelierten die Kreativitäts- und Intelligenztestergebnisse der ängstlichen Jungen positiv.

Die für Jungen und Mädchen deutlich unterschiedlichen Ergebnisse dieser Untersuchung können keinesfalls als Nachweis der Konstruktvalidität des Tests von Wallach und Kogan gewertet werden. Nur ein Teil der Ergebnisse der Jungen stützt die aus der Theorie Wallachs und Kogans abgeleiteten Untersuchungshypothesen; dem stehen die zum überwiegenden Teil stark abweichenden Ergebnisse der Mädchen gegenüber, die von Kogan und Morgan vor allem auf geschlechtsrollenspezifische Erziehungseinflüsse zurückgeführt werden.
Allerdings können die Resultate Kogans und Morgans auch nicht dazu benutzt werden, die Theorie Wallachs und Kogans zu falsifizieren oder ihr Meßverfahren abzuqualifizieren, da die "spielähnliche Testsituation" Kogans und Morgans doch einige testähnliche Züge aufwies, vor allem wegen der Gruppentestung, die offenbar völlig andere motivationale Bedingungen schuf,[1] und wegen der eventuell doch nicht unmerklichen Zeitbegrenzung. Sie machen aber auf jeden Fall deutlich, daß die Beziehungen zwischen Kreativitäts- und Intelligenztestergebnissen, Merkmalen der Testsituation, dem Geschlecht und den beiden motivationalen Variablen Schulangst und Abwehrtendenz wesentlich komplizierter sind, als Wallach und Kogan in der ersten Formulierung ihrer Theorie annahmen.

Krause (1972), der 2o2 bundesdeutsche Viertkläßler mit dem Kreativitätstest von Wallach und Kogan und einer Reihe weiterer Verfahren untersuchte, kommt zu einer Reihe von Ergebnissen, die für die Konstruktvalidität des Verfahrens bedeutsam sind.
Die positive und substantielle Beziehung zwischen Intelligenz und Kreativität wurde bereits besprochen (vgl. S. 94). In diesem Punkt dürfte eine Modifikation der Kreativitätskonzeption Wallachs und Kogans erforderlich sein: Sofern das Konstrukt Kreativität über seine Operationalisierung durch ein bestimmtes Meßinstrument hinaus Bedeutung haben soll, wird man von einer positiven und substantiellen Beziehung zu Intelligenztestergebnissen ausgehen müssen (s. dazu auch S.85).
Anzahl- und Originalitäts-Gesamtwerte korrelieren bei Krause mit .72, was bedeutet, daß durch erstere nur gut die Hälfte der Varianz

---

1) So berichten Kogan und Morgan, daß die Kinder sich in der testähnlichen Situation ruhig und diszipliniert, in der spielähnlichen Situation dagegen ziemlich ungezogen verhielten. Die Anzahl feindseliger und aggressiver Antworten im Subtest "Alternate Uses" war in dem spielähnlichen Kontext signifikant höher.

der letzteren aufgeklärt wird. Wegen des hohen Zeitaufwands, den die Originalitätsauswertung erfordert,[1] ist die Frage, welche Informationen die Originalitätsauswertung zusätzlich zu denjenigen der Auswertung nach Produktivität liefert, von erheblichem praktischen Interesse.

Krause ist der Ansicht, daß "das Originalitätsmaß eher geeignet ist, eine von der Testintelligenz unabhängige Dimension zu erfassen." (S. 113).

Mit den Ergebnissen seiner Faktorenanalyse ist die Annahme nicht zu begründen: sowohl die Produktivitäts- als auch die Originalitätsscores der 5 Untertests laden hoch auf dem "Kreativitätsfaktor" (Produktivität: $.51 \leq a \leq .76$. Md= .54, Originalität: $.47 \leq a \leq .77$, Md= .54) und sehr niedrig positiv auf dem Intelligenzfaktor (Produktivität: $.00 \leq a \leq .24$, Md= .09, Originalität: $.10 \leq a \leq .25$, Md= .17). Eine Überprüfung der Korrelationen der beiden Kreativitätsmaße mit den 7 Untertests des Leistungs-Prüf-Systems von Horn führt ebenfalls nicht zu einer Bestätigung der Hypothese Krauses: Abgesehen von den Subtests zur Erfassung des Wahrnehmungstempos und des Number-Faktors korrelieren die Originalitätswerte etwa ebenso hoch mit den LPS-Untertests wie die Anzahl-Werte. Erstaunlicherweise korreliert das Originalitätsmaß mit dem Verbalfaktor des LPS, der in erster Linie Allgemeinbildung und Rechtschreibkenntnisse erfaßt, etwas höher als der Anzahlwert,[2] das gleiche gilt für die Faktoren Wortflüssigkeit und Raumvorstellung; letztere wird durch einen Subtest erfaßt, dessen Instruktion für Viertkläßler sehr schwer zu verstehen ist.

Dieses Ergebnis wird verständlich, wenn man die Bewertung der Antworten zu der Frage nach Verwendungsmöglichkeiten für eine Zeitung, die Krause zur Illustration des Vorgehens bei der Originalitätsauswertung wiedergibt, näher betrachtet: Unter den häufig aufgetretenen Antworten, die keine Originalitätspunkte erhielten, waren u.a. "lesen", "Wissen vermehren" und "lernen". Für die nur zweimal aufgetretenen Lösungen "Erfahrung sammeln", "etwas erfahren" bzw. "neue Dinge erfahren" gab es dagegen einen Originalitätspunkt, ebenso für die nur einmal gegebene Antwort "Information holen". Hier wird deutlich, daß inhaltlich übereinstimmende Antworten je nachdem,

---

1) Sämtliche Antworten zu jedem der 39 Items des Tests werden alphabetisch aufgelistet und die Häufigkeit, mit der sie gegeben wurden, wird festgestellt. Danach werden in einem zweiten Durchgang die Originalitätspunkte für die seltenen Antworten vergeben.

2) Die Unterschiede in den Korrelationen sind nicht signifikant, sie werden hier nur interpretiert, weil in allen Fällen sprachliche Fähigkeiten involviert sind.

ob sie in umgangssprachlichen Formulierungen oder in von Kindern seltener benutzten Begriffen gegeben werden, einen Originalitätspunkt erhalten oder nicht. Es ist daher naheliegend, eine Beziehung zwischen den Originalitätsscores und dem intellektuellen oder jedenfalls sprachlichen Niveau des Elternhauses zu vermuten. Die oben erwähnten Zusammenhänge dieser Werte mit den verbalen Faktoren des LPS fänden dadurch ihre Erklärung. Aus den Angaben Krauses kann keine Bestätigung dieser Annahme abgeleitet werden. Er erwartet aufgrund ähnlicher Überlegungen eine positive Korrelation zwischen Kreativitätstestergebnissen und dem sozioökonomischen Status, findet diese Hypothese jedoch durch seine Ergebnisse nicht bestätigt. Leider kann der Feststellung "Zur Kreativität bestehen keine Beziehungen" (S. 119) nicht entnommen werden, welches Kreativitätsmaß verwendet wurde[1] und ob der Korrelationskoeffizient nicht sehr nahe bei der Signifikanzgrenze liegt. Die äußerst fragwürdige 1ostufige Einteilung der Vaterberufe von Aurin (1966), die Krause verwendete,[2] kann nicht zur Erklärung der Nullkorrelation herangezogen werden, da sich für die Intelligenztestergebnisse und einige andere Variablen der Untersuchung signifikante Rangkorrelationen ergeben haben. Somit kann die Vermutung, daß die von Krause praktizierte Originalitätsauswertung, die ohne Zusammenfassung synonymer Lösungen alle Antworten in ihrer ursprünglichen Formulierung auflistet (zweifellos das Verfahren, das die größtmögliche Auswerterobjektivität garantiert), Kinder aus Familien mit niedrigem Bildungsstand besonders stark benachteiligt, weder als bestätigt noch als widerlegt betrachtet werden.

Krause führte auch die von Wallach und Kogan verwendeten Fragebogen zur Erfassung von Schulangst und Abwehrtendenz durch. Leider sind seine Ergebnisse mit denjenigen von Wallach und Kogan nicht direkt vergleichbar, da Krause seine Daten nicht für Jungen und Mädchen getrennt und nicht varianzanalytisch, sondern korrelationsstatistisch und faktorenanalytisch verarbeitete. Dagegen untersuchte er die Anzahl-, Originalitäts-, Schulangst- und Angstabwehrwerte sowie die Beziehung der beiden motivationalen Variablen zu den Originalitäts-

---

1) Da keine Angaben gemacht werden, dürfte es sich um den durch Addition der Anzahl- und Originalitätsscores gebildeten Kreativitäts-Gesamtwert handeln. Da die Originalitätswerte immer wesentlich unter den Anzahlwerten liegen (vgl. Anmerkung 1 , S. 1o6), sind die Gesamtwerte überwiegend durch letztere determiniert.

2) Aurin vergibt an Rentner den niedrigsten (1) und an Akademiker und höhere Beamte den höchsten Rangplatz (1o). Fragwürdig sind seine Einstufungen vor allem im mittleren Bereich: Z.B. erhalten Facharbeiter und Handwerker den Rangplatz 3, Angestellte ohne besondere Ausbildung dagegen 6, selbständige Handwerker erhalten Rangplatz 4, Kaufleute (z.B. Kolonialwarenhändler) kommen auf Rangplatz 7.

werten klassenspezifisch und kam dabei zu dem überraschenden Ergebnis, daß die Mittelwerte der 6 Klassen für Produktivität, Schulangst und Abwehrtendenz signifikante Unterschiede aufwiesen und daß die Korrelationen zwischen Schulangst und Originalität für einen Teil der Klassen ebenfalls signifikant verschieden waren ($-.35 \leq r \leq .42$, Md= $-.025$). Die Abwehrtendenz-Originalitäts-Beziehungen fielen weniger unterschiedlich aus ($-.33 \leq r \leq .16$, Md= $-.05$), die Abweichungen sind nicht signifikant. Die mittleren Korrelationen lagen für beide Variablen nahe bei Null, so daß die übliche Zusammenfassung der Daten aus mehreren Klassen zu dem falschen Schluß geführt hätte, daß zwischen Originalität und Schulangst einerseits und Originalität und Abwehrtendenz andererseits keine Beziehungen bestehen.

Krause führt die Unterschiede auf die unterschiedlichen pädagogischen Strategien der Lehrer zurück; ihr jeweiliger Interaktionsstil gegenüber den Schülern bestimmt das Niveau von Schulangst und Abwehrtendenz in der Klasse und entscheidet darüber, wie die Kinder den Kreativitätstest perzipieren. Wenn er als Intelligenztest betrachtet wird, ruft er Schulangst hervor; wird er als spielähnliches und von der Schule unabhängiges Verfahren wahrgenommen, dann macht sich die hemmende Wirkung der Abwehrfunktion bemerkbar.

Die Frage, ob in einer wirklich entspannten, spielähnlichen Untersuchungssituation die von Wallach und Kogan vermutete negative Korrelation zwischen Kreativität und Abwehrtendenz auftritt, kann aufgrund der Untersuchungsergebnisse von Krause nicht beantwortet werden. Allerdings liegen indirekte Hinweise für die Richtigkeit dieser Hypothese vor: Krause legte seinen Probanden drei Fragebogen zur Familiensituation vor - zu "Strenge, Gängelung, Druck", "Vertrauensstörungen" und "äußeren familiären Schwierigkeiten". Die Stichprobe wurde nach den Kreativitätstestergebnissen medianhalbiert und die Fragebogenergebnisse der kreativen Kinder mit denjenigen der wenig kreativen verglichen. Während sich die beiden Gruppen bei der Beantwortung der Fragen zu äußeren familiären Schwierigkeiten kaum unterschieden, wurden alle übrigen Fragen, in denen eine Kritik am Elternhaus enthalten ist, von den Kreativen häufiger bejaht als von den wenig Kreativen; signifikante Unterschiede ergaben sich allerdings nur für 2 Fragen, für weitere 6 Fragen zum Elternverhalten war die negative Beurteilung der Kreativen der Tendenz nach häufiger ($.05 \leq \alpha \leq .20$).

Krause schließt aus den Ergebnissen der ad hoc zusammengestellten Skalen, daß "die Kreativen das Elternhaus vor allem in seinen psychologischen Qualifikationen negativer sehen." (S. 128). Da sich Kreative und wenig Kreative in ihren Angaben zu realen Verhaltensweisen der Eltern kaum unterschieden, kann angenommen werden, daß in den beiden anderen Bereichen auch keine objektiven Unterschiede existieren, sondern daß die Kreativen negative Urteile über ihre Eltern eher zulassen. Dies bedeutet jedoch, daß ihre Tendenz zur Verdrängung geringer ist.
Daß kreative Individuen die Beziehung zu ihren Eltern negativer sehen als wenig kreative, wird durch Untersuchungen, die andere Instrumente zur Kreativitätsmessung verwendeten, bestätigt. Kapler (197o), MacKinnon (1962) und Siegelmann (1973) fanden in den Berichten bzw. Fragebogenergebnissen Kreativer Anzeichen für eine weniger enge und warme Eltern-Kind-Beziehung als in den Kontrollgruppen, was ebenfalls als Hinweis auf eine schwächere Tendenz zur Verdrängung dieser Gruppe gewertet werden kann.

Krause führte mit sämtlichen 42 Variablen seiner Untersuchung eine Faktorenanalyse nach der Hauptachsenmethode durch: Die 6-Faktoren-Lösung erbrachte bei einer Varianzaufklärung von 44 Prozent einen deutlichen Kreativitätsfaktor, auf dem die 1o Kreativitätsmaße hoch, alle übrigen Variablen unter .3o laden,
einen Intelligenzfaktor, auf dem die Untertests des Leistungs-Prüf-Systems außer denjenigen für perceptual speed und number, die Aufsatznote, das Status-Maß, die Soziogramm-Variable "beste Ideen", das Lehrerurteil über Bravheit (mit .3o) und - mit den absolut höchsten Ladungszahlen - die Lehrerurteile zu Intelligenz (mit .75) und Ideenreichtum (mit .71) laden,
einen Schulleistungsfaktor, für den außer den Noten die 3 Lehrerurteile und die Aufsatznote (mit höherer Ladungszahl als beim Intelligenzfaktor) wichtig sind,
und einen Faktor, auf dem Schulangst und Angstabwehr mit entgegengesetzten Vorzeichen, außerdem die 3 Skalen zur elterlichen Erziehung und zu äußeren familiären Schwierigkeiten und - mit demselben Vorzeichen wie Angstabwehr - der Verbalfaktor des Leistungs-Prüf-Systems laden.

Auf dem 5. Faktor, von Krause als "Originalität verhindernde schulische Konformität" bezeichnet, laden 2 der 5 Originalitätswerte und die negativen Wahlen des Soziogramms sowie - mit entgegengesetztem Vorzeichen - die Betragensnote und das Lehrerurteil über Bravheit. Der 6. Faktor ist kaum interpretierbar; Krause bezeichnet ihn als "produktive Leistungen verhindernde Schulangst", was angesichts der Ladungszahl von .2o für Schulangst recht problematisch erscheint.

Alle übrigen Ergebnisse der Faktorenanalyse erscheinen klar interpretierbar und in Übereinstimmung mit theoretischen Erwartungen. Krauses Untersuchung erbrachte mehrere interessante Resultate. Ein wichtiges Ergebnis sind die signifikanten Mittelwertsunterschiede der Produktivitätswerte verschiedener Klassen,[1] die auf die Bedeutung der Einflußvariable Klassenlehrer hinweisen. Ein weiteres bedeutsames Resultat ist der Nachweis, daß die psychoanalytischen Konstrukte Angst[2] und Abwehrtendenz nicht ausschließlich durch frühkindliche Erfahrungen bestimmt sind, sondern auch durch Verhaltensweisen des Lehrers beeinflußt werden. Schließlich sei noch einmal darauf hingewiesen, daß Krause im Gegensatz zu Wallach und Kogan signifikante Korrelationen von mittlerer Höhe zwischen Kreativitäts- und Intelligenztestergebnissen fand und daß dieses Resultat durch andere Untersuchungen bestätigt wurde (vgl. S.94-95).

Insgesamt verdeutlichen seine Ergebnisse ein weiteres Mal die Komplexität des Untersuchungsgegenstandes, sie zeigen aber auch - trotz einiger Abweichungen von den Resultaten Wallachs und Kogans - daß der Kreativitätstest dieser beiden Autoren aufgrund seiner sorgfältigen theoretischen Fundierung ein wertvolles Forschungsinstrument ist, dessen Einsatz in theoriegeleiteten Untersuchungen Wesentliches zur Konstruktvalidierung des "problematischen Konzepts Kreativität"[3] beitragen kann.

2.3232 Übereinstimmungs- und Vorhersagevalidität

Das in den Untersuchungen zur Übereinstimmungsvalidität des Tests von Wallach und Kogan am häufigsten verwendete Außenkriterium sind

---

1) Besonders wichtig erscheint in diesem Zusammenhang, daß die Klasse mit den schlechtesten Kreativitätstestwerten in den Intelligenztestergebnissen an zweiter Stelle steht.

2) Krause untersuchte zwar nur Schulangst, aber wegen der meist recht hohen Korrelation zwischen Schulangst und allgemeiner Angst (s. dazu Wallach und Kogan 1965, S. 218-219) erscheint diese Generalisierung vertretbar.

3) Krause 1972, Untertitel

außerschulische Aktivitäten und Leistungen. Bei älteren Schülern und Studenten wird dabei häufig nach Preisen in irgendwelchen Wettbewerben, Wahlen für bestimmte Führungsaufgaben u.ä. gefragt; jüngere Schüler werden meist aufgefordert, diejenigen künstlerischen oder "wissenschaftlichen" Aktivitäten auf einer vorgegebenen Liste anzukreuzen, die sie schon aus eigenem Antrieb ausgeübt haben. Informationen dieser Art sind trotz der Gefahr,daß die Angaben durch das Bedürfnis nach sozialer Anerkennung verfälscht werden, validere Außenkriterien für einen Kreativitätstest als die Schulleistungen. Manchmal werden auch Ratings irgendwelcher kreativer Produktionen als Validitätskriterium benutzt.
Die in den Validierungsuntersuchungen zu den Torrance Tests of Creative Thinking häufig eingesetzten Lehrer- und Schülerurteile spielen dagegen kaum eine Rolle. Gelegentlich werden die Korrelationen der Kreativitätstestergebnisse mit Schulleistungen mitgeteilt. Offenbar sehr selten sind Untersuchungen, in denen versucht wird, den Valididitätsnachweis aufgrund der Korrelation des Tests von Wallach und Kogan mit einem anderen Kreativitätstest zu führen.

Die meisten Untersuchungen, die extracurriculare Aktivitäten als Außenkriterium verwendeten, wurden mit älteren Schülern oder College-Studenten durchgeführt.
Wallach und Wing (1969) untersuchten 503 College-Anfänger mit 4 Untertests des Verfahrens von Wallach und Kogan und befragten die Studenten mit einem zu diesem Zweck entwickelten Fragebogen nach außerschulischen Leistungen während der High school; außerdem verwerteten sie die Ergebnisse des Scholastic Aptitude Tests (SAT). Sie schließen aus ihren Ergebnissen:"Intelligence within the range of the present sample does not matter, but resourcefulness in producing ideas does, when it comes to talented accomplishments outside the classroom." (S. 11o). Dies gilt sowohl für ein allgemeines Kriterium außerschulischer Leistungen als auch für die nach Bereichen getrennten Ergebnisse: Die nach dem Test von Wallach und Kogan Hochkreativen (oberes Drittel der Verteilung) konnten auf mehr Leistungen in bildender Kunst, Literatur, Wissenschaft und auf mehr Führungsaktivitäten verweisen als die wenig Kreativen (unteres Drittel). Dieses Ergebnis gilt für beide Geschlechter. Die Unterschiede zwischen den beiden Gruppen waren im allgemeinen ausgeprägter, wenn die Gruppenbildung aufgrund der Produktivitäts-

werte vorgenommen wurde, nur für schriftstellerische Aktivitäten war der Zusammenhang mit den Originalitätswerten enger.
Keine Unterschiede zwischen Hochkreativen und wenig Kreativen ergaben sich auf dem Gebiet sozialer Aktivitäten, für Schauspiel und Musik, ein durchaus plausibles Ergebnis, da sich in diesen Bereichen seltener innovative Fähigkeiten entfalten können.
Die Untersuchung von Wallach und Wing hat einige methodische Schwächen, die die innere und äußere Validität der Ergebnisse einschränken: Die Stichprobe bestand aus Freiwilligen, die nur 4o Prozent der College-Anfänger der für die Untersuchung ausgewählten Universität darstellten. Da es sich sowohl bei der Testbearbeitung als auch bei den außerschulischen Leistungen um Aktivitäten handelt, die ohne äußeren Zwang oder Druck ausgeübt wurden, sieht Nicholls (197o) in der Tendenz, in Situationen ohne äußere Kontrolle aufgabenrelevante Motivationen zu entwickeln, die Variable, die zu den von Wallach und Wing festgestellten Zusammenhängen führte.
Die SAT-Mittelwerte waren sehr hoch, es handelte sich also um eine Gruppe Hochintelligenter. Die Korrelationen zwischen den Variablen der Untersuchung werden nicht mitgeteilt, obwohl auch in dieser Stichprobe mit eingeschränkter Intelligenzstreuung die Möglichkeit einer positiven Beziehung zwischen Kreativitätstest- und SAT-Ergebnissen nicht a priori ausgeschlossen werden kann. Eine Reanalyse der Daten durch Murphy (1973) zeigt allerdings, daß nur eines der 8 Subtestergebnisse signifikant, aber sehr niedrig mit dem verbalen SAT-Score korrelierte.

Daß eine so aufwendige Untersuchung die erhobenen Daten nur auf Mittelwertsunterschiede zwischen 2 Extremgruppen überprüft, ist ein äußerst fragwürdiges Vorgehen: Ein Drittel der vorhandenen Informationen bleibt ungenutzt, die Stärke der Zusammenhänge unbekannt und eventuelle kurvilineare Beziehungen bleiben unentdeckt; außerdem befinden sich unter den Ergebnissen der vielen t-Tests (etwa 8o) Signifikanzen, die zufällig zustande kamen.
So können die Ergebnisse der Arbeit von Wallach und Wing nur als Validitätshinweis gewertet werden, obwohl sie bei einer adäquaten Methode der Datenverarbeitung eventuell als überzeugender Nachweis der Gültigkeit des Verfahrens von Wallach und Kogan gelten könnten.

Wagner (1972) untersuchte die Senioren zweier High-School-Klassen mit dem Kreativitätstest von Wallach und Kogan, einem Intelligenztest und dem Fragebogen von Wallach zur Erfassung außerschulischer Leistungen. Der Kreativitätstest wurde z.T. mit und z.T. ohne Zeitbegrenzung durchgeführt.
Unter beiden Bedingungen waren Intelligenz- und Kreativitätstestergebnisse unkorreliert, und die Beziehungen zwischen den Kreativitätstestergebnissen und dem Fragebogen zur Erfassung extracurricularer Betätigungen waren signifikant positiv.
Rotter, Langland und Berger (1971) führten eine Validitätsuntersuchung mit 61 Zweitkläßlern durch, bei der ein von Torrance entwickelter Fragebogen nach kreativen Aktivitäten als Außenkriterium verwendet wurde.
Während die Ergebnisse des Intelligenztests weder mit denjenigen des Verfahrens von Wallach und Kogan noch mit dem Validitätskriterium korrelierten, ergaben sich mehrere signifikante Korrelationen zwischen den Kreativitätstestergebnissen und den Fragebogenbereichen "creative writing and composition", "writing and reading", "dramatics" und "life sciences", die Korrelationen waren allerdings recht niedrig ($.15 \leq r \leq .37$, Md =. 28). Die im Gegensatz zu der Untersuchung von Wallach und Wing signifikante Beziehung der Testergebnisse mit "dramatics" erklärt sich dadurch, daß die Zweitkläßler nicht in erster Linie nach darstellerischen Leistungen gefragt wurden, sondern danach, ob sie sich selbst ein Stück ausgedacht hätten u.ä..
Werts (1967) versuchte zwar zu zeigen, daß auch die außerschulischen Leistungen während der High School durch die Schulnoten vorhergesagt werden können, wurde aber von Holland und Richards (1967) widerlegt. Die beiden Autoren verweisen auf mehrere Untersuchungen, in denen die Noten der High School gute Prädiktoren der College-Noten, nicht jedoch der außerhalb des Studiums liegenden Leistungen sind, während extracurriculare Aktivitäten während der High-School-Zeit sowohl die Studienleistungen als auch die nicht direkt auf das Studium bezogene Produktivität vorhersagen. Ihre Ergebnisse werfen ein günstiges Licht auf das Validitätskriterium "extracurriculare Aktivitäten", das auf dem Kontinuum Testverhalten - reales Verhalten doch relativ nahe dem Pol des realen Verhaltens lokalisiert zu sein scheint und das offenbar eine recht gute zeitliche

Stabilität besitzt. Es ist bedauerlich, daß Validierungsuntersuchungen mit diesem Kriterium in der derzeitigen Schulsituation wenig Erfolg versprechen, weil die Anzahl der Schüler, die noch Zeit und Kraft zu außerschulischen produktiven Betätigungen finden, im Zeichen des numerus clausus und der Lehr- und Arbeitsstellenknappheit zu gering sein dürfte.
Eine Untersuchung, in der die Ratings kreativer Produkte als Validitätskriterium verwendet wurden, wurde von Bartlett und Davis (1974) durchgeführt. Sie untersuchten 37 studentische Teilnehmer eines Kreativitäts-Kurses mit dem Test von Wallach und Kogan und beurteilten die künstlerischen oder schriftstellerischen Produkte, Erfindungen oder Vorschläge für kreative Unterrichtsmethoden, die einen Teil der Kursanforderungen bildeten, auf einer 7stufigen Skala nach dem Kreativitätsgrad. Die Korrelationen der Ratings mit den Produktivitäts-, Originalitäts- und Kreativitäts-Gesamtwerten betrugen .33, .33 und .35. Obwohl signifikant, sind diese Korrelationen doch sehr niedrig, so daß das positive Urteil von Bartlett und Davis über diesen "Validitätsnachweis" nicht gerechtfertigt erscheint.

Eine Validierung ihres Tests mit Hilfe der objektiv erfaßten oder von den Lehrern beurteilten Schulleistungen wurde von Wallach und Kogan nie in Betracht gezogen. Trotzdem ist die Frage, welche Beziehungen - bei kontrollierter Intelligenz - zwischen den Testergebnissen und den Schulleistungen bestehen, nicht uninteressant, zumal hierzu Ergebnisse amerikanischer und deutscher Untersuchungen vorliegen.
Wagner (1972) fand für die ohne Zeitbegrenzung ermittelten Kreativitätstestergebnisse signifikante Beziehungen zum High School grade-point average. Murphy (1973) verschickte den Test von Wallach und Kogan während der Sommerferien an 371 Schüler einer High School und erhielt 14o beantwortete Testhefte zurück. Die Schule stellte Intelligenztestergebnisse und Noten zur Verfügung. Eine multiple Regressionsanalyse erbrachte folgende Ergebnisse: Die Intelligenzmaße korrelierten mit .69 mit den Noten, die Kreativitätstestergebnisse mit .22, beide Prädiktoren klärten zusammen etwa 5o Prozent der Notenvarianz auf (R =.71). Die Produkt-Moment-Korrelationen der Prädiktorvariablen mit dem Kriterium unterschieden sich fast nicht von den entsprechenden Partialkorrelationen, was auf die Unabhängigkeit der

beiden Variablen hinweist. Die Erhöhung der Korrelation durch Hinzunahme des zweiten Prädiktors war statistisch signifikant, aber so gering, daß die zeitaufwendige Durchführung und vor allem Auswertung des Verfahrens kaum lohnend erscheint.
Bei Krause (1972) korrelierten das Intelligenz- und das Kreativitätsmaß mit .44. Der Zusammenhang zwischen Kreativität und Schulleistung war auch nach Auspartialisierung der Intelligenz für die sprachlichen Noten signifikant, die Korrelation sinkt durch die Auspartialisierung von .29 auf .17. Die Kreativitäts-Gesamtwerte, die überwiegend durch die Produktivitätsscores determiniert sind,[1)] korrelieren meist fast ebenso hoch mit den Noten wie die Intelligenztestergebnisse; für Mitarbeit, Rechnen, Heimatkunde und "bildhaftes Gestalten" liegen die Korrelationen zwischen .36 und .41. Die Originalitätswerte korrelieren insgesamt - außer in Aufsatz, Musik und Leibesübungen - deutlich niedriger mit den Noten.
Krauses Ergebnisse werden durch diejenigen von Grote, Hajek u.a. (1969) bestätigt, die ebenfalls signifikante, meist sogar noch engere Zusammenhänge zwischen den Kreativitätstestwerten und den Noten in bildnerischem Gestalten, Musik, Rechnen und Lesen errechneten.
Aufgrund dieser Ergebnisse wird man auch für das Verfahren von Wallach und Kogan von einer signifikanten, aber niedrigen Korrelation zwischen Kreativitätstestwerten und Noten ausgehen müssen; bei zusätzlicher Durchführung eines Intelligenztestes tragen die Kreativitätstestergebnisse zwar signifikant, aber unerheblich zur Verbesserung der Vorhersage bei.

Richards (1974) führte eine Untersuchung durch, in der 394 Marinerekruten 3 semantische Tests der divergenten Produktion von Guilford und 3 Subtests des Verfahrens von Wallach und Kogan (allerdings in einer auf je 2 Items reduzierten Fassung) bearbeiteten. Obwohl die Stichprobe aus jungen Erwachsenen bestand, sollen die Ergebnisse doch kurz beschrieben werden, weil keine weitere Untersuchung gefunden werden konnte, in der eine andere Kreativitätsoperationalisierung als Validitätskriterium für das Verfahren von Wallach und Kogan verwendet wurde.

---

1) S. dazu Anmerkung 1), S. 107

Weitere Vorteile der Arbeit sind, daß die Intelligenzstreuung der Stichprobe nicht eingeschränkt war und daß der Test von Wallach und Kogan auf eine Art und Weise durchgeführt wurde, die auch in Untersuchungen mit größeren Probandenzahlen realisierbar ist, nämlich im Gruppenrahmen und mit großzügig bemessener Zeitbegrenzung.
Die aus je 2 Items bestehenden "Subtests" von Wallach und Kogan interkorrelierten mit .48, die 3 Guilford-Untertests mit .52; nach einer Korrektur wegen Unreliabilität stiegen die gemittelten Korrelationen auf .69 bzw. .7o. Die mittlere Korrelation der 3 Guilford- mit den 3 Wallach-Kogan-Aufgaben betrug .46 (korrigiert .66). Da die Stichprobe ziemlich groß und - jedenfalls im Hinblick auf die Intelligenz - in ihrer Streuung nicht eingeschränkt war, wird man bei Erwachsenen mit diesen relativ niedrigen Validitätskoeffizienten rechnen müssen, wenn ein Kreativitätstest als Außenkriterium verwendet wird.[1)]
Insgesamt betrachtet haben die Untersuchungen zur Übereinstimmungsvalidität zu weniger günstigen Ergebnissen geführt als diejenigen zur Konstruktvalidität. Dies geht jedoch wahrscheinlich weniger zu Lasten des Verfahrens von Wallach und Kogan als zu Lasten der Schwierigkeit, Außenkriterien für einen Kreativitätstest zu finden.

Kogan und Pankove (1971) führten eine Untersuchung zur zeitlichen Stabilität und zur prädiktiven Validität der Ergebnisse des Tests von Wallach und Kogan durch.
Die beiden Stichproben von 29 und 72 Fünftkläßlern[2)] kamen aus Mittelschichtfamilien und besuchten 2 verschiedene Schulen. Bei der ersten Testung wurden 2 Untertests - Alternate Uses und Pattern Mea-

---

1) Da sich die Kreativitätskonzeptionen Guilfords und Wallachs und Kogans grundsätzlich unterscheiden - Guilford geht von multiplen unabhängigen Faktoren mit niedrigen positiven Korrelationen zu Faktoren der konvergenten Produktion und anderen Intelligenzfaktoren aus, Wallach und Kogan erwarten eine einheitliche, intelligenzunabhängige Kreativitätsdimension - sind die Schlußfolgerungen, die Richards aus ihren Ergebnissen zieht, ebenfalls von Interesse: "Results for all creative thinking tests came closer to the Wallach-Kogan unifactorial view of associative creative ability than the Guilford multifactorial view of divergent production, whether judging from mean correlations or the factor analysis." (S.14) Allerdings zeigten nicht nur die Guilford-Tests, sondern auch die Aufgaben von Wallach und Kogan eine positive Beziehung zu den Intelligenztestergebnissen (Guilford-Test: r =.26, korrigiert .32, Wallach-Kogan-Aufgaben: r =.17, korrigiert .21).

2) Die Stichproben waren bei den ersten Untersuchungen größer, aber hier werden nur die Probanden berücksichtigt, die auch an der zweiten Untersuchung teilnahmen.

nings - im Einzelversuch und ohne Zeitbegrenzung durchgeführt. Nach 5 Jahren wurden die Schüler erneut untersucht, u.a. auch mit dem von Wallach und Wing entwickelten Fragebogen zur Erfassung außerschulischer Aktivitäten und Leistungen. Die multiple Regressionsanalyse zur Vorhersage der Fragebogenergebnisse erbrachte folgende Ergebnisse für die kleinere der beiden Schulen, aus der die Stichprobe von 29 Kindern kam: Die einfache Korrelation der Produktivitätswerte mit dem Kriterium betrug .366;[1] bei dieser Stärke der Beziehung werden gut 13 Prozent der Kriteriumsvarianz durch den Prädiktor aufgeklärt (der entsprechende F-Wert ist nicht signifikant, seine Wahrscheinlichkeit ist aber kleiner als .1o). Durch Hinzunahme der Originalitätswerte stieg der Prozentsatz der aufgeklärten Varianz kaum an (von 13,4 auf 13,8 Prozent!). Erst durch den Intelligenzquotienten der 5. Klasse kam es zu einem signifikanten Anstieg der multiplen Korrelation auf .575, was allerdings auch nur eine Aufklärung der Kriteriumsvarianz von 33 Prozent bedeutet.
Die überaus unbefriedigende prädiktive Validität der beiden Subtests von Wallach und Kogan ist sicher in erster Linie auf die zeitliche Instabilität des Prädiktorverhaltens zurückzuführen (s. dazu S.89-91); wenn nämlich die Produktivitätswerte der zum Zeitpunkt der zweiten Untersuchung durchgeführten Kreativitätstestaufgaben als vierter Prädiktor in die Regressionsanalyse eingegeben werden, steigt der Prozentsatz der aufgeklärten Kriteriumsvarianz noch einmal hochsignifikant an (von 33,o auf 49,6 Prozent).
In der größeren der beiden Schulen besaßen die beiden Kreativitätsmaße überhaupt keine prädiktive Validität; der Anteil der durch beide Maße aufgeklärten Kriteriumsvarianz betrug o,1 Prozent.

Eine Beurteilung der Untersuchungsergebnisse zur Übereinstimmungs- und Vorhersagevalidität kann nur zu dem Resultat kommen, daß die Korrelationen der Kreativitätstestwerte mit den bisher verwendeten Außenkriterien fast durchweg unbefriedigend niedrig sind. Dies ist eventuell in erster Linie darauf zurückzuführen, daß es bisher noch nicht gelungen ist, valide Außenkriterien für einen Kreativitätstest zu entwickeln. In den wenigen Fällen, in denen ein Kriterium gewählt wurde, das als einigermaßen repräsentativ für reales kreatives Verhalten betrachtet werden kann - wie z.B. die aus eigenem Antrieb erbrachten extracurricularen Leistungen -, verminderte eventuell das Bedürfnis, sich sozial erwünscht darzustellen, oder ähnliche,

---

1) Kogan und Pankove geben zwar einen Koeffizienten von .41 an, aber der in Tabelle 4 mitgeteilte Determinationskoeffizient von .134 widerspricht diesem Wert (vgl. Kogan und Pankove, S. 21).

die Ergebnisse von Fragebogen verfälschende Tendenzen die an sich höheren Validitätskoeffizienten. Eine abschließende Beurteilung der Validität des Tests von Wallach und Kogan kann erst erfolgen, wenn Untersuchungen mit besseren Außenkriterien vorliegen. Auch wenn die ersten Arbeiten zur Konstruktvalidierung überwiegend ermutigende Resultate erbrachten, erscheint für ein als "Kreativitätstest" bezeichnetes Verfahren der Nachweis eines mehr als mäßigen Zusammenhangs mit realem kreativem Verhalten oder mit anderen validen Prädiktoren kreativer Leistungen unabdingbar.

2.33 Ausgewählte Untersuchungsergebnisse

Aus der großen Zahl der mit dem Verfahren von Wallach und Kogan durchgeführten Untersuchungen wurde wieder eine Reihe von Arbeiten zu den Auswirkungen unterschiedlicher Testbedingungen, zur Auswertung, zu Geschlechts- und Schichtabhängigkeit und zu den Beziehungen zwischen den Kreativitätstestergebnissen, Intelligenz und Schulleistung ausgewählt.
Mit Hilfe der Ergebnisse dieser Untersuchungen sollten die Schlußfolgerungen überprüft werden, die aus den Resultaten der mit dem Torrance-Test durchgeführten Arbeiten gezogen wurden (vgl. S. 76-77). Außerdem sollte versucht werden, die bisher offen gebliebene Frage nach der optimalen Testsituation aufgrund der mit dem Test von Wallach und Kogan erzielten Untersuchungsergebnisse zu beantworten (s. dazu 55-60).

2.331 Auswirkungen unterschiedlicher Testbedingungen

Die Frage der Auswirkungen verschiedener Untersuchungssituationen auf die Testergebnisse wurde bereits mehrfach gestreift, aber noch nicht im Zusammenhang besprochen.

Obwohl Wallach und Kogan (1965) die Ansicht vertreten, daß nur eine Individualtestung in entspannter, spielähnlicher Atmosphäre und ohne Zeitbegrenzung zu intern konsistenten und intelligenzunabhängigen Kreativitätstestergebnissen führen kann, wurden in vielen Untersuchungen modifizierte und z.T. auch völlig abweichende Testbedingungen angewendet. Tabelle 4 gibt einen Überblick über Untersuchungen, die das Verfahren von Wallach und Kogan zumindest teilweise durchführten und die Testsituation ausreichend charakterisieren.

Überblick über die Auswirkungen verschiedener Testsituationen auf die Ergebnisse von Wallach-Kogan-Aufgaben

Tabelle 4: Überblick über die Auswirkungen verschiedener Testsituationen auf die Ergebnisse von Wallach-Kogan-Aufgaben

Autor(en)	Stichprobe	Testbedingung(en)	Ergebnisse	Übereinstimmung mit den Hypothesen von Wallach und Kogan?
Williams und Fleming 1969	36 Vierjährige	Individualtestung unter spiel- und testähnlichen Bedingungen	Signifikante Interkorrelationen der Kreativitätsaufgaben unter beiden Bedingungen, insignifikante Korrelationen zwischen Kreativitäts- und Intelligenztestergebnissen unter beiden Bedingungen	? Da Vierjährige vermutlich die negativen Konsequenzen von Test- und Prüfungssituationen noch nicht kennengelernt haben, kann die Frage nicht beantwortet werden.
Ward 1969b	53 4-bis 5jährige hochintelligente Kinder	Individualtestung ohne Zeitbegrenzung und mit liebevoller, anerkennender Behandlung der Kinder ——— Testung in anregungsarmer und -reicher Umgebung	Keine Korrelation zwischen Intelligenz- und Kreativitätstestergebnissen und z. T. auch Nullkorrelationen zwischen den Kreativitätstestscores ——— Kreative Kinder gaben im Gegensatz zu unkreativen mehr Antworten in einer anregungsreichen Umgebung als in einer anregungsarmen.	Das letztgenannte Ergebnis widerspricht den Resultaten von Wallach und Kogan, die offenbar nicht auf sehr viel jüngere Kinder generalisiert werden können. ——— Die Überlegenheit der kreativen Kinder in der Benutzung aufgabenrelevanter Cues aus der Umgebung - verstanden als Offenheit der Wahrnehmung - kann als indirekte Bestätigung der Hypothesen von Wallach und Kogan betrachtet werden.
Elkind, Deblinger und Adler 1970	8 5-bis 6-, 8 7-bis 8-, 8 9-bis 10- und 8 11-bis 12jährige Schüler der "World of Inquiry School" in New York	Jedes Kind wurde 2 Mal - mit äquivalenten Hälften dreier Subtests von Wallach und Kogan - getestet, einmal im Anschluß an eine "interessante" und einmal im Anschluß an eine "uninteressante" Tätigkeit; die Kinder wußten, daß sie nach der Testung zu der vorangegangenen Tätigkeit zurückkehren würden. Effekte der Reihenfolge waren ausbalanciert.	Kinder, die nach einer langweiligen Tätigkeit getestet wurden, gaben hochsignifikant mehr und eine hochsignifikant höhere Anzahl origineller Antworten. Dieses Resultat gilt für alle untersuchten Altersstufen und alle 3 Subtests. Die Reihenfolge der Testbedingungen hatte keinen Effekt.	? Wallach und Kogan erörtern solche meist zufallsbedingten Einflüsse auf die aufgabenrelevante Motivation ihrer Probanden nicht; evtl. halten sie ihr Meßinstrument nicht für so anfällig gegenüber kurzzeitig wirksamen Beeinflussungen. Außerdem ist es möglich, daß die relative Position der Kinder in beiden Testsituationen gleich oder nahezu gleich war.
Ward 1968	34 7-bis 8jährige Jungen	Individualtestung in entspannter Atmosphäre und ohne Zeitbegrenzung	Hohe interne Konsistenz der Kreativitätsmaße, keine Zusammenhänge mit den Intelligenztestergebnissen	Ja
Rotter, Langland und Berger 1971	61 Zweitkläßler	Individualtestung in spielähnlicher Atmosphäre ohne Zeitbegrenzung	Hohe Korrelation der Gesamtwerte der verbalen und der visuellen Untertests, keine Zusammenhänge mit den Intelligenztestergebnissen. Korrelationen mit 40 Bereichen der "Checklist of Creative Activities" von Torrance zwischen .15 und .37	Ja ? ?
Krause 1972	202 Vierkläßler	Gruppentestung in relativ entspannter Atmosphäre ohne Zeitbegrenzung	Die Interkorrelationen der 10 Kreativitätsmaße sind etwa von gleicher Höhe wie dieselben Werte bei Wallach und Kogan; die Korrelation zwischen Kreativitäts- und Intelligenztestergebnissen ist .44.	Nein (Gruppentestung) Ja

Tabelle 4: Fortsetzung

Autor(en)	Stichprobe	Testbedingung(en)	Ergebnisse	Übereinstimmung mit den Hypothesen von Wallach und Kogan?
Grote, Hajek u.a. 1969	25 Viertkläßler	Gruppentestung in relativ entspannter Atmosphäre, den Kindern standen 2 volle Schulstunden zur Verfügung, die jedoch von niemand voll ausgenutzt wurden.	Hohe interne Konsistenz der Kreativitätsmaße (höher als bei Wallach und Kogan); Median der Rangkorrelationen von .3o für die Kreativitäts-Intelligenz-Beziehung	Nein (Gruppentestung) Ja
Ward, Kogan und Pankove 1972	191 Fünftkläßler aus Familien mit relativ niedrigem sozioökonomischem Status	Individualtestung mit Lob und Ermutigung und ohne Zeitbegrenzung. Jedes Kind wurde in 2 Sitzungen mit je einem verbalen und einem visuellen Subtest untersucht (ausbalancierte Reihenfolge innerhalb und zwischen den Sitzungen). Während des 1. Untertests der 2. Sitzung erhielt ein Drittel der Kinder einen Cent für jede Antwort, dem 2. Drittel wurde dieselbe Belohnung für das Ende der Testdurchführung versprochen, das letzte Drittel war die Kontrollgruppe. Beim 2. Untertest wurden auf dieselbe Weise "gute Einfälle" unmittelbar oder nach Beendigung der Testung belohnt.	Hohe interne Konsistenz der Anzahlwerte, keine Beziehung zu den Intelligenztestergebnissen. Sowohl die unmittelbare als auch die hinausgeschobene Belohnung führten zu hochsignifikanten Erhöhungen der Anzahlwerte, nicht jedoch zu signifikanten Qualitätsverbesserungen. Die Manipulation der motivationalen Bedingungen ließ die relative Position der Pbn in den Testergebnissen nahezu unverändert.	Ja (s.dazu auch unter Punkt 2.334) Zu den Ergebnissen des Experiments ist kein Urteil möglich; da die individuellen Unterschiede bestehen blieben, dürfte kein Widerspruch zu den theoretischen Erwartungen Wallachs und Kogans vorliegen.
Kogan und Morgan 1969	1o9 Fünftkläßler aus Familien der unteren Mittelschicht	Gruppentestung; für die Experimentalgruppe schulnahe streßinduzierende Testsituation mit betonter Zeitbegrenzung, für die Kontrollgruppe nicht schulleistungsrelevante entspannte Testsituation mit "unmerklicher" Zeitbegrenzung (wenn ein Schüler sagte, er sei mit einer der Aufgaben nicht fertig, wurde er auf eine spätere Möglichkeit zur Weiterarbeit verwiesen, die aber in Wirklichkeit nicht existierte).	Höhere Anzahl- und Originalitätswerte in dem Subtest "Alternate Uses" und z.T. für den Subtest "Pattern Meanings" in der testähnlichen Untersuchungssituation. Die Interkorrelationen der Anzahl- und Originalitätswerte der beiden Aufgabengruppen waren in der testähnlichen Situation höher als in der entspannten. Für Mädchen ergaben sich in der spielähnlichen Situation keine Beziehungen zwischen Kreativitäts- und Intelligenztestergebnissen, bei den Jungen korrelierte der Originalitätswert von "Alternate Uses" .42 mit dem IQ. In der testähnlichen Situation korrelierten Intelligenz- und Kreativitätstestergebnisse nicht.	Nein Nein Überwiegend Ja Nein
Nicholls 1971 (die Ergebnisse der Untersuchung werden z.T. vollständiger referiert von Vernon 1971)	229 Fünftkläßler	114 Schüler wurden mit 2 Parallelformen aus Wallach-Kogan-Aufgaben getestet, zuerst einzeln und unter denselben Bedingungen wie bei Wallach und Kogan; nach Schulleistungs- u. Intelligenztests wurde die Parallelform unter testähnlichen Bedingungen mit Zeitbegrenzung durchgeführt. Eine Kontrollgruppe erhielt dieselbe Form nur ein Mal in einem testähnlichen Kontext.	Hohe Interkorrelationen der Kreativitätsmaße für die spielähnliche Testung, geringere für die testähnliche Untersuchung. Nullkorrelationen der Kreativitätsmaße mit Intelligenztestergebnissen in der spielähnlichen, signifikante Intelligenz-Kreativitäts-Beziehungen in der testähnlichen Untersuchungssituation.	Ja Ja

Tabelle 4: Fortsetzung

Autor(en)	Stichprobe	Testbedingung(en)	Ergebnisse	Übereinstimmung mit den Hypothesen von Wallach und Kogan?
Leith 1972	106 9-,11- und 13jährige Schüler aus Konstanz	Testähnliche und entspannte Untersuchungssituationen, im ersten Fall mit betonter, im zweiten Fall mit unauffälliger Zeitbegrenzung; in beiden Fällen Gruppentestung.	Höhere Anzahl- und Originalitätswerte und größere Streuungen für die testähnliche Untersuchungssituation.	Nein
			Die Interkorrelationen der Kreativitätsmaße (Kendall's W) sind für die testähnliche Situation etwas, aber nicht signifikant höher (evtl. nur eine Folge der größeren Streuung, s.o.).	?
			Keine signifikanten Intelligenz-Kreativitäts-Korrelationen für beide Testbedingungen.	Nein (Gruppentestung)
Kogan und Pankove 1971	Schule A: 46 Fünft- bzw. Zehntkläßler	Schule A: Individualtestung in entspannter Atmosphäre und ohne Zeitbegrenzung bei der ersten Untersuchung in der 5. und bei der zweiten Untersuchung in der 10. Klasse	Schule A: Hohe interne Konsistenz der Produktivitätswerte der Jungen und Mädchen für die erste und 2. Testung, die Originalitätswerte der Jungen zeigten in der 2. Testung z.T. Nullkorrelationen mit anderen Kreativitätsmaßen und untereinander.	Ja für die Produktivitätswerte, Nein für die Originalitätswerte der Jungen in der 2. Testung
			Keine signifikanten Beziehungen zum IQ in der ersten Testung, einige signifikant positive Korrelationen für Jungen zwischen Produktivitätsscores der 2. Testung und dem IQ.	Ja für die 1. Testung, Nein für die Produktivitätswerte der Jungen in der 2. Testung
	Schule B: 116 Fünftkläßler (bei der Retestung in der 10. Klasse: 72). Die Schüler beider Schulen kamen aus Familien mit vergleichbarer sozioökonomischer Position.	Schule B: Bei der ersten Testung wie in Schule A, in der 10. Klasse Gruppentestung - getrennt für Jungen und Mädchen - in entspannter Atmosphäre; den Schülern wurde gesagt, daß die Untersuchung nicht schulleistungsrelevant sei. Keine eindeutige Aussage bezüglich einer Zeitbegrenzung bei der 2. Testung.	Die Retest-Reliabilitäten (bei Verwendung derselben Aufgaben) waren für Jungen und Mädchen fast ausnahmslos nicht signifikant.	? (Wiederholung der gleichen Aufgaben)
			Die zeitliche Stabilität der Kreativitätsmaße (bei Verwendung anderer Aufgaben) war für Jungen ebenfalls sehr schlecht, für Mädchen etwas besser.	Nein
			Signifikant positive, aber sehr niedrige Korrelationen mit außerschulischen Leistungen in der 10. Klasse. Schule B: Hohe interne Konsistenz der Kreativitätsmaße der Jungen und Mädchen für die erste Testung, keine Korrelationen mit dem IQ der 5. Klasse.	? Ja
			Fast durchweg signifikante und z.T. sehr hohe Interkorrelationen der Kreativitätsmaße von Jungen und Mädchen bei der 2. Testung.	Nein (Gruppentestung)
			Überwiegend signifikante, aber mäßige Retest-Reliabilitäten (bei gleichen Aufgaben) für Jungen, ausschließlich insignifikante Retestreliabilitäten zwischen -.11 und .27 für Mädchen.	?
			Überwiegend signifikante, aber relativ niedrige Koeffizienten für die zeitliche Stabilität der Kreativitätsmaße (andere Aufgaben) für Jungen, Nullkorrelationen für Mädchen.	Diese Frage kann wegen des Unterschieds in den Bedingungen der 1. und 2. Testung nicht beantwortet werden.

Tabelle 4: Fortsetzung

Autor(en)	Stich-probe	Testbedingung(en)	Ergebnisse	Übereinstimmung mit den Hypothesen von Wallach und Kogan?
			Für Jungen signifikante und hochsignifikante Korrelationen der Kreativitätsmaße mit dem IQ der 1o. Klasse, keine Zusammenhänge für die Mädchen.	Ja für Jungen (Gruppentestung)
			Überhaupt keine Zusammenhänge zwischen den Kreativitätsmaßen der 1. und 2. Testung und den außerschulischen Leistungen der 1o. Klasse. (Dieses Validitätskriterium korrelierte auch nicht mit dem IQ!)	Nein
Vernon 1971	14 achte Klassen aus 2 High Schools	Gruppentestung mit testähnlicher Untersuchungsbedingung und Zeitbegrenzung versus Gruppentestung mit entspannter, "spielähnlicher" Atmosphäre ohne Zeitbegrenzung der Aufgaben, allerdings war die Gesamt-Testzeit wie bei der testähnlichen Untersuchung auf 2 Schulstunden begrenzt.	Z.T. höhere Kreativitätswerte der unter "spielähnlichen" Bedingungen getesteten Gruppe, vor allem bei den Originalitätswerten.	Ja
			Diese Gruppe hatte auch höhere Interkorrelationen der Kreativitätstestergebnisse als die in einer testähnlichen Situation untersuchten Schüler,	Ja
			und ihre Kreativitätswerte korrelierten höher mit Validitätskriterien.	Ja
			Die Kreativitäts-Intelligenz-Korrelationen waren für beide Gruppen etwa gleich und signifikant positiv.	? (Gruppentestung, aber mit unterschiedlicher Testatmosphäre)
Murphy 1973	14o High School Sophomores	Brieftestung, d.h. die Schüler bearbeiteten die Tests freiwillig, zu Hause, zu einem selbstgewählten Zeitpunkt und ohne Zeitbegrenzung	Hohe interne Konsistenz der Kreativitätsmaße, kaum signifikante Korrelationen mit verschiedenen Intelligenztestergebnissen (Md=.o8).	Ja
			Signifikante, aber sehr niedrige Korrelation mit Zensuren.	?
Wagner 1972	2 Klassen High School Seniors	Gruppentestung mit und ohne Zeitbegrenzung	Die Reliabilität der Subtests war höher bei der zeitlich unbegrenzten Testung.	Ja
			Der Zusammenhang der Testergebnisse mit Intelligenztestergebnissen wurde durch die Variable Testzeit nicht beeinflußt.	Nein
			Auch die Beziehung der Produktivitätswerte zu extracurricularen Leistungen war für beide Testbedingungen signifikant positiv.	?
Wallach und Wing 1969	5o3 College-Anfänger	Freiwillige Teilnahme an einer Gruppentestung ohne Zeitbegrenzung.	Hohe interne Konsistenz der Kreativitätsmaße, keine Beziehung zu Intelligenztestergebnissen.	? Hier stellt sich die Frage nach dem Äquivalent einer "spielähnlichen Testsituation" für Erwachsene.

Die in Tabelle 4 zusammengestellten Ergebnisse bieten ein überaus verwirrendes Bild. Die Beurteilung, ob sie mit den Hypothesen von Wallach und Kogan übereinstimmen, war z.T. nicht möglich, da diese beiden Autoren keine Aussagen darüber machen, welche Komponenten ihrer Untersuchungssituation sie für bedeutsam halten - die Individualtestung, die entspannte, schulferne Atmosphäre, den Wegfall der Zeitbegrenzung oder die mündliche Beantwortung der Testitems - oder ob sie alle 4 Bedingungen als gleich wichtige Voraussetzungen für die Erfassung einer einheitlichen intelligenzunabhängigen Kreativitätsdimension betrachten.

Daß diese Hypothese falsch wäre, geht aus Tabelle 4 klar hervor: zumindest eine einheitliche Kreativitätsdimension mit hoher interner Konsistenz der Kreativitätsmaße wird auch durch Gruppentestungen erfaßt (z. B. von Krause 1972, Grote, Hajek u.a. 1969, Vernon 1971, Murphy 1973, Wallach und Wing 1969, Cropley und Maslany 1969, Richards 1974).

Eine Individualtestung ist für ältere Jungen evtl. sogar leistungsmindernd: Die Resultate von Kogan und Pankove (1971) für die Jungen aus Schule A zeigen einen signifikanten Abfall der Kreativitätstestergebnisse von der 5. zur 1o. Klasse; die Nullkorrelationen zwischen den Ergebnissen der ersten und zweiten Testung für die Jungen dieser Schule deuten in dieselbe Richtung (vgl. Tabelle 4). Die Mädchen der Schule A haben zwar ebenfalls insignifikante Retestreliabilitäten, aber die in der 1o. Klasse neu eingesetzten Kreativitätstestaufgaben korrelieren überwiegend signifikant und z.T. recht hoch mit den Testergebnissen der 5. Klasse. Kogan und Pankove interpretieren die Interaktion zwischen Testsituation und Geschlecht dahingehend, daß ältere Jungen sich in der anonymen Gruppentestung, wie sie in Schule B durchgeführt wurde, wohler fühlen, während für Mädchen die individuelle Testsituation günstiger ist. Wegen der sehr kleinen Probandenzahlen bedarf diese Interpretation jedoch weiterer Überprüfungen.

Während Gruppentestungen also nach diesen Ergebnissen die interne Konsistenz der Kreativitätstestergebnisse nicht zu beeinträchtigen scheinen, liegen mehrere Anzeichen dafür vor, daß die Kreativitätswerte, die im Gruppenversuch erhoben wurden, signifikant mit In-

telligenztestergebnissen korrelieren (s.z.B. die Untersuchungen von Krause 1972, Grote, Hajek u.a. 1969, Kogan und Pankove 1971, Vernon 1971, Richards 1974); zu diesem Punkt liegen allerdings auch widersprüchliche Ergebnisse vor (z.B. von Kogan und Morgan 1969, Murphy 1973, Wallach und Wing 1969).

Die Art der Beantwortung der Items durch die Probanden kann in der Gruppenteststituation nicht variiert werden und konnte daher auch nicht auf ihre Auswirkungen hin überprüft werden.

Auch die Manipulation der Variablen Testzeit ist in Gruppenuntersuchungen vor allem mit jüngeren Probanden recht schwierig; besonders die Kinder, die langsamer arbeiten, aber auch diejenigen, die sich gerne lange mit den Aufgaben beschäftigen würden, dürften durch die Beobachtung, daß andere Kinder ihre Testhefte abgeben, unter Druck gesetzt werden.

Trotzdem wurden auch Gruppenuntersuchungen ohne Zeitbegrenzung durchgeführt, z.B. von Krause 1972, Grote, Hajek u.a. 1969, Wagner 1972, Wallach und Wing 1969, Cropley und Maslany 1969. Ein Vergleich der Ergebnisse mit denjenigen von Untersuchungen, in denen vergleichbare Gruppentestungen mit Zeitbegrenzung stattfanden, erbringt kein eindeutiges Resultat: In Untersuchungen ohne Zeitbegrenzung wurden hohe interne Konsistenzen der Kreativitätsmaße gefunden und sowohl signifikante Beziehungen zu Intelligenztestergebnissen (z.B. von Grote, Hajek u.a. 1969 und von Krause 1972) als auch Nullkorrelationen (z.B. von Wagner 1972, von Wallach und Wing 1969 und von Cropley und Maslany 1969).
Aber auch Gruppenuntersuchungen mit Zeitbegrenzung erbrachten hohe Interkorrelationen der Kreativitätstestergebnisse (z.B. Kogan und Morgan 1969, Richards 1974) und ebenfalls sowohl signifikante Kreativitäts-Intelligenz-Zusammenhänge (Nicholls 1971, Vernon 1971, Richards 1974) als auch Nullkorrelationen (Kogan und Morgan 1969, Leith 1972, Wagner 1972).

Eventuell ist der Zeitfaktor allein doch keine so wichtige Variable wie es nach der Theorie von Wallach und Kogan zu erwarten wäre; die Ergebnisse von Ward (1969 a) deuten ebenfalls in diese Richtung (vgl. S. 98-99). Die widersprüchlichen Ergebnisse der Untersuchungen ohne Zeitbegrenzung könnten eventuell damit erklärt werden, daß in

Gruppenuntersuchungen mit Viertkläßlern ein Zeitdruck für die länger Arbeitenden entsteht, wenn nach und nach ein Teil der Kinder aufhört zu schreiben, während sich Studenten in Testsituationen dieser Art sicher relativ unabhängig verhalten.
Das Problem der Nullkorrelationen zwischen Intelligenz und Kreativität in Untersuchungen mit Zeitbegrenzung bleibt allerdings bestehen. Eine Überprüfung der Auswirkungen der Atmosphäre während der Testung auf die interne Konsistenz der Kreativitätstestwerte und ihre Beziehung zu Intelligenztestergebnissen trägt nicht zur Klärung bei, sondern verwirrt das widerspruchsvolle Bild noch mehr: Zwei Untersuchungen (Kogan und Morgan 1969 und Leith 1972) fanden höhere interne Konsistenzen der Kreativitätsmaße in der testähnlichen, stressinduzierenden Untersuchungssituation; in einer anderen Arbeit (Vernon 1971) war die interne Konsistenz für die in einer spielähnlichen, entspannten Atmosphäre ermittelten Ergebnisse höher. Ebenso wurden in testähnlichen Untersuchungssituationen sowohl signifikante Kreativitäts-Intelligenz-Zusammenhänge (Nicholls 1971, Vernon 1971) als auch Nullkorrelationen gefunden (Kogan und Morgan 1969, Leith 1972). Für "spielähnliche" Untersuchungen ergaben sich dieselben Widersprüche: Leith fand keine, Vernon und Kogan und Morgan fanden signifikante Korrelationen zwischen Kreativitäts- und Intelligenztestergebnissen, letztere allerdings nur für Jungen und nur für eines der Kreativitätsmaße.

Bei der Beurteilung dieser Untersuchungsergebnisse ist zu berücksichtigen, daß die Dichotomie - spielähnliche versus testähnliche Untersuchungssituation - sehr wenig konkret ist und dementsprechend sehr unterschiedlich operationalisiert werden kann.
Außerdem erhebt sich die Frage, wieso das Kontinuum spielähnliche, entspannte bis prüfungsähnliche, angsterregende Testsituation fast immer auf diese Dichotomie reduziert wurde.[1] Die Frage nach der günstigsten Untersuchungsbedingung kann eventuell nur beantwortet werden, wenn außer den beiden extremen auch eine mittlere, einen mäßigen Druck zur intensiven Bearbeitung des Tests erzeugende Situation in den experimentellen Plan aufgenommen wird. (Eventuell

1) Eine Ausnahme ist die Untersuchung von Laux und Zinn (1974), in der jedoch nicht der Kreativitätstest von Wallach und Kogan verwendet wurde.

nimmt die spielähnliche Individualtestung von Wallach und Kogan eine solche Mittelstellung hinsichtlich des Drucks zur Aufgabenbearbeitung ein; die z.T. ungünstigen Ergebnisse der "spielähnlichen" Gruppenuntersuchungen wären dann durch die zu geringe Motivation der Probanden für die Testbearbeitung zu erklären.)
Die entspannten bzw. testähnlichen Untersuchungssituationen der verschiedenen Arbeiten sind sicher nicht vergleichbar, aber selbst wenn es sich um objektiv gleiche Bedingungen handeln würde, könnten sie von den Probanden sehr verschieden erlebt worden sein; die Ergebnisse von Krause (1972) zeigen die Berechtigung der Annahme, daß gerade bei Testdurchführungen in Schulklassen durch objektiv gleiche oder zumindest sehr ähnliche äußere Bedingungen von Klasse zu Klasse unterschiedliche motivationale Voraussetzungen geschaffen werden können (vgl. S. 1o7-1o8).
Abgesehen von dem Interaktionsstil des Klassenlehrers wirken sich jedoch auch das Geschlecht und bestimmte Persönlichkeitsmerkmale als Moderatorvariablen auf die kreative Produktion aus; neben den häufiger untersuchten Dimensionen Angst und Abwehrtendenz hat eventuell auch die Extraversion einen Einfluß (s. dazu Leith 1972). Alle diese Variablen interagieren vermutlich in komplexer Weise mit verschiedenen Aspekten der Testsituation, was bedeutet, daß es einer großen Zahl systematischer experimenteller Untersuchungen bedarf, bevor sichere Aussagen darüber möglich sind, welche Testsituation die validesten Kreativitätstestergebnisse erbringt. Vermutlich wird diese Frage je nach Validitätskriterium unterschiedlich beantwortet werden müssen; schon für die sehr häufig verwendeten, aber wenig befriedigenden Validitätskriterien der in Tabelle 4 berücksichtigten Untersuchungen - hohe interne Konsistenz, keine oder zumindest sehr niedrige Korrelationen mit Intelligenztestwerten - fielen die Ergebnisse oft unterschiedlich aus. Die Frage der prädiktiven Validität wurde bisher selten, diejenige nach der Zahl irrelevanter Antworten, die nach Laux und Zinn (1974) ebenfalls von der Testsituation abhängt, wurde in Untersuchungen mit dem Test von Wallach und Kogan bisher offenbar noch gar nicht berücksichtigt.
Daß eines der beiden in allen Arbeiten berücksichtigten Kriterien, die Unabhängigkeit der Kreativitätsmaße von Intelligenztestergebnissen, problematisch ist, wurde bereits erörtert (vgl.S. 85).

Anhand der vorliegenden Untersuchungsergebnisse kann die Frage, welche Testsituation für die Erfassung einer intelligenzunabhängigen Kreativitätsdimension am günstigsten ist,auch gar nicht beantwortet werden, da die Intelligenz mit den unterschiedlichsten Verfahren gemessen wurde.
Für das Kriterium der internen Konsistenz der Kreativitätstestwerte scheint es nach den vorliegenden Untersuchungsergebnissen nicht wichtig zu sein, das Verfahren von Wallach und Kogan in Einzeltestung und mit mündlicher Aufgabenbeantwortung durchzuführen; auch eine großzügige bemessene Zeitbegrenzung dürfte die interne Konsistenz nicht oder kaum beeinträchtigen.

Zur Test- oder Spielähnlichkeit der Untersuchungssituation kann für Schüler mittlerer Klassenstufen[1)] nur folgende Hypothese aufgestellt werden: Eine schulferne und damit nicht angsterregende Testsituation dürfte günstig sein, wenn es gelingt, das Interesse der Kinder für die Testaufgaben zu wecken. Vor allem bei klassenweisen Testungen besteht eventuell die Gefahr, daß die Kinder bei Wegfall der schulüblichen extrinsischen Motivierung durch Angst vor schlechten Noten keine Lust haben, sich mit dem Test zu beschäftigen, der wie Klassenarbeiten die schriftliche Beantwortung von Fragen erfordert. Die schlechteren Ergebnisse der in "spielähnlichen" Gruppentestungen untersuchten Schüler sind vermutlich so zu erklären (gefunden in den Untersuchungen von Kogan und Morgan 1969, von Leith 1972 und von Laux und Zinn 1974 mit einem anderen Kreativitätstest).

Vernon (1971) macht aufgrund eines Literaturüberblicks und der in seiner eigenen Untersuchung gesammelten Erfahrungen einige Vorschläge zur optimalen Gestaltung der Testsituation bei Gruppenuntersuchungen in Schulen, die auch nach den hier zusätzlich berücksichtigten Untersuchungen vertretbar erscheinen:

a) In der Instruktion sollte betont werden, daß das kein Test im üblichen Sinn ist.
b) Die Aufgabe der Probanden sollte klar definiert werden, etwa in dem Sinn, daß es keine "richtigen" Antworten gibt, sondern daß möglichst viele unübliche,einfallsreiche, originelle Lösungen gefunden werden sollen.

---

1) Eine Generalisierung der Empfehlungen Wallachs und Kogans auf höhere Altersstufen erscheint gerade im Hinblick auf die Untersuchungssituation besonders schwierig; wie müßte das Äquivalent eines spielähnlichen Kontextes für ältere Schüler und für Studenten aussehen?

c) Den Probanden sollte klar gemacht werden, daß von ihnen erwartet wird, daß sie sich auf die Testbearbeitung konzentrieren.
d) Der Testleiter sollte kein Lehrer sein und möglichst nichts mit der Schule zu tun haben. Er soll sich freundlich, entspannt und ermutigend verhalten und doch die Kontrolle über die Situation nie aus der Hand geben.
e) Die Probanden sollten den Test möglichst in ihrem eigenen Tempo bearbeiten können, wobei Vorkehrungen getroffen werden müssen, damit diejenigen, die fertig sind, sich mit Lesen o.ä. beschäftigen können, ohne die anderen Schüler zu stören.[1)]

2.332 Testauswertung

Sowohl das Problem der theoretischen Konzeption des Zusammenhangs zwischen Anzahl- und Originalitätswerten als auch die dazu vorliegenden Untersuchungsergebnisse wurden bereits angesprochen (s. dazu S. 85-86 und S.97-98). Tabelle 5 zeigt noch einmal die Ergebnisse verschiedener Autoren im Überblick.

Die Koeffizienten der einzelnen Arbeiten sind sehr unterschiedlich, wobei keine Gesetzmäßigkeit wie z.B. Altersabhängigkeit, Einfluß der Begabungsstreuung o.ä. vorzuliegen scheint.[2)] Da Wallach und Kogan keine detaillierten Anweisungen für die Originalitätsauswertung gegeben haben, ist es möglich, daß diese bei den verschiedenen Untersuchern sehr unterschiedlich gehandhabt wurde,[3)] was die verschieden hohen Korrelationen mit den Anzahlwerten erklären könnte.
In den meisten Untersuchungen korrelieren die Anzahlwerte höher mit irgendwelchen Validitätskriterien als die Originalitätswerte, was eine Folge der geringeren Reliabilität der letzteren sein kann.[4)]

---

1) Zweifellos ist dieses Verfahren, das eine Konsequenz der assoziativen Kreativitätstheorie von Wallach und Kogan berücksichtigt, optimal. Nach den hier zusätzlich besprochenen Ergebnissen ist es allerdings wahrscheinlich, daß die Validität der Testergebnisse durch eine großzügig bemessene Zeitbegrenzung nicht beeinträchtigt wird (vgl. S.125-126).

2) Die besonders niedrigen Koeffizienten von Wallach und Kogan sind allerdings auf ihre besondere Art der Berechnung der Korrelationen zurückzuführen, vgl. Anmerkung 1, S. 91

3) S. dazu z.B. S. 106-107

4) Wallach und Kogan (1965) fanden für die Originalitätswerte der 5 Subtests bedeutend geringere Split-half-Reliabilitäten (nach Spearman-Brown) als für die Anzahlwerte; ebenso korrelierte jeder Subtest-Originalitätswert niedriger mit den 9 anderen Kreativitätsmaßen als der Anzahlwert desselben Untertests (vgl. Wallach und Kogan, S. 41, Tab. 1 u.S. 46, Tab. 7); Richards (1974) erhielt für die Originalitätswerte von Guilford und von Wallach und Kogan die niedrigsten Reliabilitätskoeffizienten.

Tabelle 5: Anzahl-Originalitäts-Korrelationen verschiedener Untersuchungen

Autor(en)	Stichprobe	Untertests	Koeffizienten
Wallach und Kogan 1965	151 Fünftkläßler	Instances Alternate Uses Similarities Pattern Meanings Line Meanings	.o8 .67 .71 .29 .64
Krause 1972	2o2 Viertkläßler	Instances Alternate Uses Similarities Pattern Meanings Line Meanings	.63 .62 .57 .71 .54
Grote, Hajek u.a. 1969	25 Viertkläßler	Instances Alternate Uses Pattern Meanings Line Meanings	.88 .73, .83 .73, .97 .92, .93 Koeffizienten für je 1 Item
Ward 1968	41 5jährige Jungen, 46 5jährige Mädchen		Jungen / Mädchen
		Instances Alternate Uses Pattern Meanings	.89, .92 .95, ".85 .84, ".82
	34 8jährige Jungen	Instances Alternate Uses Pattern Meanings	.92 .94 .98
Murphy 1973	14o High school Sophomores	Alternate Uses Similarities Pattern Meanings	.8o .76 .86
Wallach und Wing 1969	5o3 College-Anfänger	Alternate Uses Similarities Pattern Meanings Line Meanings	.55 .58 .66 .63
Cropley 1968	124 australische Studenten im ersten Jahr	Instances Alternate Uses Similarities Pattern Meanings Line Meanings	.61 .6o .61 .61 .63

Dies beruht vermutlich darauf, daß die zur Untersuchung der Zusammenhänge meist verwendete Produkt-Moment-Korrelation für die Originalitätswerte nicht das geeignete Verfahren ist, da diese offenbar selten oder gar nicht normal verteilt auftreten.

Krause (1972) erhält sowohl für seine Gesamtstichprobe als auch für die einzelnen Klassen bei einigermaßen normalverteilten Produktivitätsscores bimodale Verteilungen für die Originalitätswerte. Er schließt daraus: "Wir müssen also annehmen, daß es zwei voneinander verschiedene Gruppen gibt." (S. 1o1). Die Folgerung erscheint etwas verfrüht, zumal andere Autoren andere Verteilungen fanden; Richards (1974), deren Stichprobe aus 394 Marinerekruten bestand, erhielt Originalitäts-Verteilungen, deren Modalwert bei Null lag. Bhavnani und Hutt (1972) untersuchten 12o 7-bis 8jährigen Jungen und Mädchen mit 4 der 5 Subtests von Wallach und Kogan und erhielten bei nicht allzu stark von der Normalverteilung abweichenden Anzahlwerten für Jungen und Mädchen Originalitätsverteilungen, die für den verbalen Teil des Tests den Modalwert bei Null, für die visuellen Aufgaben den Modalwert bei 2 hatten. Bei Zusammenfassung der Originalitätswerte aller 4 Subtests lag der Modalwert bei Null, und die Verteilungen für Jungen und Mädchen fielen relativ gleichmäßig - ohne einen zweiten Modalwert - ab.

Die Operationalisierung der Originalität als statistische Seltenheit der Lösungen ist auch unabhängig von der Verteilungsform nicht unproblematisch: der Auswerter muß einerseits vermeiden, für völlig abwegige und bizarre Antworten Originalitätspunkte zu vergeben, andererseits läuft er dauernd Gefahr, Zusammenhänge, die der Proband zwischen der jeweiligen Aufgabe und seiner Lösung gesehen oder hergestellt hat, nicht zu erkennen und dadurch originelle Antworten nicht zu honorieren. Hinzu kommt, daß eine Antwort zu einem bestimmten Zeitpunkt oder in einer bestimmten Stichprobe originell sein kann, kurze Zeit später - etwa nach den ersten Sendungen einer Fernseh-Serie - oder in einer anderen Probandengruppe, eventuell sogar schon in der Parallelklasse, kann derselbe Begriff mehrmals auftreten. Das bedeutet, daß es nicht möglich ist, "Originalitäts-Normen" zu erstellen, die dann für weitere Stichproben verwendet werden könnten.

Wenn jedoch die Vergabe der Originalitätspunkte in jeder Untersuchung neu festgelegt wird, entsteht ein Problem, auf das bereits Krause (1972) hingewiesen hat: Je nach Stichprobengröße ist es leichter oder schwieriger, Originalitätspunkte zu erhalten; eine Antwort, die in einer Gruppe von 3o Probanden nur einmal aufgetreten ist, ist statistisch weniger selten als eine Antwort, die in einer Stichprobe von 3oo Personen einzigartig war.

Vernon (1971) schlägt für Untersuchungen mit größeren Probandenzahlen eine Modifikation der Originalitätsauswertung vor, die das Problem der Abhängigkeit der Schwierigkeit der Originalitätspunkte vom Stichprobenumfang löst und außerdem eine Verbesserung der Verteilungsform der Originalitätsscores bewirkt:
Aus der Gesamtstichprobe werden nach dem Zufall 1oo Probanden ausgewählt, deren Item-Beantwortungen alphabetisch aufgelistet werden. Für jede Antwort wird ausgezählt, wie oft sie in dieser Zufallsstichprobe gegeben wurde. Antworten, die 15 Mal oder häufiger aufgetreten sind, erhalten 0 Punkte und werden in einer Art Normenliste festgehalten. Lösungen, die 6 bis 14 Mal vorkamen, werden mit einem Originalitätspunkt bewertet und ebenfalls in die Normenliste aufgenommen. Antworten, die nur 1 bis 5 Mal gegeben wurden, erhalten 2 Originalitätspunkte; sie brauchen nicht in der Normenliste vermerkt zu werden, da alle Lösungen, die dort nicht aufgeführt sind - was bedeutet, daß sie in der "Norm-Stichprobe" entweder gar nicht oder höchstens bis zu 5 Mal aufgetreten sind - 2 Originalitätspunkte erhalten.
Mit Hilfe der Normenliste werden dann alle Antworten der Gesamtstichprobe, die im ersten Auswertungsdurchgang einen Produktivitätspunkt erhielten, nach ihrer Originalität bewertet.
Es ist klar, daß bei dieser Art der Auswertung Mittelwert und Streuung der Originalitätsscores größer werden, die Verteilung symmetrischer wird und sich damit der Normalverteilung annähert.
Trotz der Vorteile dieses Originalitätsmaßes bleibt die grundsätzliche Schwierigkeit der subjektiven Entscheidung, wann eine Antwort noch eine eindeutige Beziehung zu dem Kreativitätstestitem aufweist und wann nicht mehr, bestehen. Vernon ist der Ansicht, daß subjektive Entscheidungen dann getroffen werden können, wenn sie festgehalten und konsequent auf alle Probanden der Stichprobe angewendet werden.
Die dargestellten Probleme und Schwierigkeiten der Originalitätsauswertung machen es erforderlich, den Informationsgewinn zu überprüfen, den sie erbringt; es muß also untersucht werden, ob die Originalitätsscores zu bestimmten Validitätskriterien andere - aber nicht einfach nur niedrigere - Korrelationen haben als die Produktivitätswerte. Diese Untersuchungen müssen jedoch mit einer ver-

besserten Originalitätsauswertung durchgeführt werden; die bisher ganz überwiegend weniger engen Zusammenhänge der Originalitätsscores mit anderen Kreativitätstestergebnissen und mit Außenkriterien sind sehr wahrscheinlich sowohl durch die geringere Streuung als auch durch die für die Berechnung von Produkt-Moment-Korrelationen ungeeignete Verteilungsform der Originalitätswerte bedingt.[1)]

Bisher scheint für Kreativitätsaufgaben von Wallach und Kogan nur eine Untersuchung mit einer verbesserten Originalitätsauswertung vorzuliegen, diejenige von Vernon (1971). Vernon vergleicht die von ihm vorgeschlagene Originalitätsauswertung mit derjenigen nach der Anzahl der Antworten und einem aus beiden berechneten Index, dem Prozentsatz origineller Antworten. Im Hinblick auf die interne Konsistenz erbringen die nach Produktivität ausgewerteten Aufgaben verschiedener Autoren [2)] die besten Ergebnisse; mit einer Reihe von anderen Variablen, die allerdings nur z.T. als Validitätskriterien betrachtet werden können, korrelieren dagegen durchweg die Originalitätswerte am höchsten. Daß letztere nicht nur mit Interesse für Kunst, Lehrerurteilen über die Neugierde der Schüler und Rorschachergebnissen substantiell korrelieren, sondern auch mit Intelligenz- und Wortschatztestergebnissen,erschwert ein Urteil darüber, welches Maß vorzuziehen ist. Die Korrelation der Anzahl- mit den Originalitätsscores betrug .76 und ist damit so hoch, daß sich diese Frage eventuell erübrigt.
Krause (1972) verwendete die von Wallach und Kogan vorgeschlagene Originalitätsauswertung, nur vergibt er mit Hinweis auf seine um über 3o Prozent größere Stichprobe nicht nur für einmalige, sondern auch

---

1) Grote, Hajek u.a. (1969) berechneten für die Daten ihrer 25 Pbn Rangkorrelationen. Sowohl bei den Interkorrelationen der Kreativitätsmaße als auch bei den Korrelationen mit Außenkriterien ergaben sich überwiegend, aber mit einer Reihe von Ausnahmen, höhere Koeffizienten für die Anzahlwerte. Die Erklärung dieser Ausnahmen bereitet Schwierigkeiten: Da bei einem N von 25 auch der Schwierigkeitsgrad der Originalitätswerte stark reduziert ist, kann nicht beurteilt werden, ob die relativ günstigen Ergebnisse für die Originalitätsscores auf die Erhöhung der Streuung, auf die Verwendung des adäquaten Korrelationskoeffizienten oder auf beides zurückzuführen sind.

2) Vernon (1971) verwendete in seiner Untersuchung neben Items aus Wallach-Kogan-Aufgaben auch Subtests von Torrance und 2 Aufgaben von Thurstone.

für 2 Mal gegebene Antworten einen Originalitätspunkt. Trotz der bimodalen Verteilung der Originalitätswerte erhält er für diese engere Zusammenhänge mit einigen Untertests des LPS als für die Anzahlwerte (die Unterschiede sind allerdings nicht signifikant). Die hierzu vorgeschlagene Interpretation (s. dazu S. 1o6-1o7). widerspricht derjenigen von Krause, kann jedoch mit Hilfe der zur Verfügung stehenden Informationen nicht abgesichert werden. Im Fall einer Bestätigung würde sie gegen die Verwendung von Originalitätsscores sprechen, zumindest wenn sie wie bei Krause ermittelt wurden.
Allerdings enthält die Arbeit von Krause eventuell andere Hinweise auf die diagnostische Valenz der Originalitätswerte. Z.B. werden die Beziehungen zwischen Schulangst und Originalität und zwischen Abwehrtendenz und Originalität mitgeteilt, nicht jedoch diejenigen zwischen den beiden motivationalen Variablen und Produktivität. Falls Krause hier im Gegensatz zu seinem sonstigen Verfahren, entweder die Korrelationen für beide Kreativitätsmaße oder diejenige für den Kreativitäts-Gesamtwert anzugeben, den Originalitätsscore bevorzugte, weil er enger mit Schulangst und Abwehrtendenz zusammenhängt, spräche das für eine Berücksichtigung oder jedenfalls weitere Überprüfung des Originalitätsmaßes. In dieselbe Richtung deutet ein Ergebnis im Zusammenhang mit den Lehrerurteilen über Intelligenz, Ideenreichtum und Bravheit der Schüler: alle 3 Kreativitätsvariablen korrelieren hochsignifikant positiv mit den 3 Lehrerratings, nur der Originalitätsscore korreliert mit der Bravheitsschätzung insignifikant.

Ein weiterer Hinweis auf eine eventuelle diagnostische Valenz der Originalitätsscores wird den Untersuchungsergebnissen zu den Beziehungen zwischen Geschwisterkonstellation und Kreativität entnommen: Krause stellte fest, daß die Jüngsten in der Geschwisterreihe die höchsten Produktivitäts-, Originalitäts- und Kreativitäts-Gesamtwerte hatten, der Unterschied zwischen den Jüngsten und den zwischen Ältesten und Jüngsten geborenen Kindern ist für Produktivität und Gesamtwert auf dem 5%-Niveau gesichert, für das Originalitätsmaß ist er sogar hochsignifikant.
Nach der Interpretation Krauses sind die Jüngsten eventuell deshalb am kreativsten, weil die Ängstlichkeit der Mütter und die daraus resultierende Überprotektion mit jedem weiteren Kind abnehmen und

weil die soziale Umwelt von Kindern mit älteren Geschwistern stimulierend und im Hinblick auf Rollenangebote sehr abwechslungsreich ist.

Schließlich erscheint auch noch ein Ergebnis von Bhavnani und Hutt (1972), die 12o Sieben- bis Neunjährige mit 4 Subtests von Wallach und Kogan untersuchten, von Interesse: Während die 6o Mädchen hochsignifikant höhere Anzahlwerte als die 6o Jungen hatten, ergaben sich keine geschlechtsspezifischen Unterschiede für die Originalitätsscores.

Obwohl jeder einzelne dieser Hinweise kaum von der Unerläßlichkeit der Originalitätsauswertung überzeugen kann, lassen sie es in ihrer Gesamtheit doch angezeigt erscheinen, die Frage, ob das Originalitätsmaß Informationen liefert, die in den Produktivitätswerten nicht enthalten sind, weiter zu überprüfen.

2.333 Geschlechtsabhängigkeit der Ergebnisse des Kreativitätstests von Wallach und Kogan

In der Untersuchung von Wallach und Kogan fielen nur Anzahl- und Originalitätswert des Subtests "Instances" für Jungen und Mädchen unterschiedlich aus; Jungen hatten signifikant höhere Anzahl- und hochsignifikant höhere Originalitätsscores. Im Kreativitätsgesamtwert - bei Wallach und Kogan die Summe der z-Werte aller 1o Kreativitäts-Subtestergebnisse - unterschieden sich Jungen und Mädchen nicht.
Nach Kogan (1974) wurden in den meisten Untersuchungen, die das Geschlecht der Probanden berücksichtigten, keine Unterschiede in den Produktivitätswerten der Jungen und Mädchen gefunden.

Ein Vergleich der Interkorrelationen der 1o Kreativitätsmaße bei Jungen und Mädchen führt allerdings zu dem Ergebnis, daß die Leistungen der letzteren konsistenter sind: 44 der 45 Korrelationskoeffizienten der Mädchen waren hochsignifikant, aber nur 27 Koeffizienten der Jungen (6 weitere Koeffizienten waren mit einer Zufallswahrscheinlichkeit von 5 % signifikant). Eventuell kann dies als Hinweis auf eine allgemein höhere und somit weniger von der jeweiligen Aufgabe abhängige Anstrengungsbereitschaft der Mädchen gewertet werden.

Obwohl Wallach und Kogan für eine ganze Reihe von Variablen, die mit den Kreativitätstestscores korrelierten, geschlechtsspezifische

Unterschiede fanden, die durch andere Untersuchungen bestätigt wurden (z.B. von Kogan und Morgan 1969, Kogan und Pankove 1971), wird das Geschlecht in vielen Untersuchungen nicht berücksichtigt (u.a. in der Arbeit von Krause 1972).

Eine Untersuchung, deren zentrale Fragestellung die geschlechtsspezifischen Unterschiede in Kreativitätstestergebnissen ist, wurde von Bhavnani und Hutt (1972) mit 4 der 5 Subtests von Wallach und Kogan durchgeführt.[1] Die Stichprobe bestand aus je 60 sieben- bis neunjährigen Jungen und Mädchen. Die Mädchen hatten sowohl in den beiden verbalen Subtests als auch in den beiden Untertests mit visuellen Aufgabenstimuli signifikant höhere Produktivitätsscores als die Jungen, der Unterschied für letztere war noch ausgeprägter als für erstere.
Auch bei den Originalitätswerten waren die Leistungen der Jungen und Mädchen in den verbalen Aufgaben ähnlicher als in den visuellen; allerdings waren die Differenzen weder für die verbalen oder visuellen Scores noch für die Gesamtwerte aus beiden signifikant.
Bhavnani und Hutt weisen darauf hin, daß es bei der in vielen Untersuchungen festgestellten höheren Wortflüssigkeit der Mädchen dieser Altersstufe erstaunlich gewesen wäre, wenn sie zu verbalen Assoziationsaufgaben nicht mehr Lösungen angeboten hätten als die Jungen. Das Originalitätsmaß hänge dagegen weniger von sprachlichen Fähigkeiten ab und daher seien hier keine geschlechtsspezifischen Unterschiede aufgetreten. Die Autoren vermuten, daß zu einem späteren Zeitpunkt, wenn sich die sprachlichen Fähigkeiten der Jungen denjenigen der Mädchen angeglichen haben, die ersteren in Kreativitätstests oder zumindest in den Originalitätswerten dieser Verfahren besser abschneiden. Diese Vermutung wird gestützt durch den Hinweis auf die Untersuchung von Wallach und Kogan, die für ihre ältere Stichprobe Ergebnisse erhielten, nach denen die Jungen - allerdings nicht signifikant - besser waren als die Mädchen.

Der Befund von Bhavnani und Hutt, daß Mädchen höhere Anzahlwerte in "Kreativitätsaufgaben" erreichen als Jungen, wird durch einige Unter-

---

1) Daß bei dieser Fragestellung ausgerechnet der einzige Untertest weggelassen wurde, in dem sich Jungen und Mädchen der Stichprobe von Wallach und Kogan signifikant unterschieden, dürfte sich auch nicht durch den Hinweis auf die geringere Split-half-Reliabilität (immerhin .51 bei 4 Items, alle anderen Subtests haben mindestens doppelt so viele Items!) und seine niedrigere Korrelation mit den übrigen Untertests begründen lassen.

suchungen mit den Torrance-Tests bestätigt (vgl. S. 66-67); ihre Interpretation erscheint allerdings zumindest in 2 Punkten angreifbar. Einmal können die Fähigkeiten, die durch den Test von Wallach und Kogan erfaßt werden, sicher nicht mit sprachlichen Fähigkeiten gleichgesetzt werden. Nach Guilfords Modell wären letztere dem symbolischen, erstere dem semantischen Inhaltsbereich zuzuordnen. Sprachliche Fähigkeiten spielen bei der Beantwortung der Aufgaben von Wallach und Kogan nur insofern eine Rolle, als es vorkommen kann, daß ein Kind einen Einfall hat, ihn aber nicht äußert, weil es ihn nicht verbalisieren kann. Ob der größere Wortschatz oder die größere Wortflüssigkeit der Mädchen ausreicht, um ihre besseren Testergebnisse zu erklären, muß bezweifelt werden. Eine alternative Erklärung wäre die größere Anstrengungsbereitschaft der Mädchen in schulischen oder als schulähnlich erlebten Situationen (Bhavnani und Hutt testeten die Kinder in Fünfer- und Sechsergruppen); sie müßte zumindest zusätzlich berücksichtigt werden.

Außerdem erscheint die Hypothese, daß die Jungen infolge ihres raschen Aufholens in den sprachlichen Fähigkeiten bereits mit 1o bis 11 Jahren zum Teil bessere Kreativitätstestergebnisse haben als die Mädchen, wenig fundiert. Der von den Ergebnissen Bhavnanis und Hutts abweichende Befund von Wallach und Kogan könnte auch mit Besonderheiten der beiden Stichproben erklärt werden; schließlich handelt es sich in beiden Fällen nicht um Zufallsstichproben und um relativ kleine Probandengruppen. Einen Hinweis in dieser Richtung gibt eine Gegenüberstellung der Subtestergebnisse der Wechsler Intelligence Scale (WISC) für Jungen und Mädchen bei Wallach und Kogan: Die Jungen erreichen signifikant höhere Werte im Wortschatz-Untertest als die Mädchen, ein auf dieser Altersstufe ungewöhnliches Ergebnis.

Neuere Untersuchungen mit deutschsprachigen Stichproben stellen überhaupt keine oder nur sehr geringfügige geschlechtsspezifische Unterschiede in den sprachlichen Fähigkeiten von Jungen und Mädchen fest.[1)]

Die Ergebnisse und Hypothesen von Bhavnani und Hutt werden auch durch amerikanische und englische Untersuchungen nur teilweise bestätigt.

---

1) Z.B. Gastel 1964, Grimm 1973, Kasielke 1967, Oevermann 197o

Kogan und Morgan (1969) fanden für ihre 1o- bis 11jährigen Jungen höhere Originalitätsscores in den beiden Subtests Alternate Uses und Pattern Meanings als für Mädchen, die Differenz war aber nur für die letztgenannte Aufgabengruppe signifikant.
Nach den wenigen Untersuchungen, die das Geschlecht der Probanden berücksichtigen, verläuft die weitere Entwicklung nicht so, wie es Bhavnani und Hutt erwarten. Entweder wurden keine geschlechtsabhängigen Unterschiede in den Ergebnissen gefunden (z.B. von Cropley und Maslany 1969 für Studenten, von Ward und Warren 1971 für 283 Erwachsene, die an einer Art Originalitäts-Wettbewerb teilnahmen[1]) oder die Kreativitätstestwerte der Jungen entwickelten sich nicht in der vermuteten Richtung. Kogan und Pankove (1971), die 2 relativ kleine Stichproben in der 5. Klasse und 5 Jahre später erneut untersuchten, stellten weder bei Einzel- noch bei Gruppentestung eine signifikante Steigerung der Testergebnisse der Jungen fest. In der Gruppe, mit der auch bei der 2. Testung Einzeluntersuchungen durchgeführt wurden, stiegen die Leistungen der Mädchen insignifikant an, während diejenigen der Jungen fast signifikant abfielen ($p < .1o$); in der Schule, deren Schüler in Gruppen untersucht wurden, stiegen nur die Leistungen der Mädchen in einem der beiden Subtests hochsignifikant an, alle anderen Veränderungen waren nicht signifikant.

Da die Geschlechtsabhängigkeit von Kreativitätstestergebnissen bisher noch nicht unter Einbeziehung der Altersvariablen untersucht wurde, sind noch keine definitiven Aussagen möglich. Bei dem derzeitigen Stand der Forschung erscheint eine Überprüfung von Kreativitätstestergebnissen auf geschlechtsspezifische Unterschiede unbedingt erforderlich.

2.334 Schichtabhängigkeit der Ergebnisse des Kreativitätstests von Wallach und Kogan

Die Frage der Schichtabhängigkeit des Verfahrens von Wallach und Kogan ist noch nicht systematisch untersucht worden, so daß auch hierzu nur vereinzelte Hinweise zusammengestellt werden können.

---

1) Eine Rundfunkstation in New York forderte ihre Hörer auf, humorvolle und originelle "kleine grüne Dinge" einzusenden.

Die Untersuchung von Wallach und Kogan wurde mit Kindern aus Familien der mittleren bis oberen Mittelschicht durchgeführt, ebenso eine Reihe der nachfolgenden Untersuchungen. Kogan und Morgan (1969) scheinen die ersten gewesen zu sein, die zwei der 5 Untertests des Verfahrens von Wallach und Kogan mit Kindern der unteren Mittelschicht durchführten. Da die beiden Aufgabengruppen jedoch mit Zeitbegrenzung und als Gruppentest durchgeführt wurden, können Mittelwerte und Standardabweichungen nicht mit den von den Testautoren mitgeteilten Werten verglichen werden.

Krause (1972), der den Test ohne Zeitbegrenzung, aber im Gruppenrahmen durchführte, teilt Mittelwerte und Standardabweichungen mit, die ganz erheblich unter denjenigen von Wallach und Kogan liegen. Seine Probanden stammen aus Familien mit niedrigerem sozioökonomischen Status, da sie jedoch durchschnittlich ein knappes Jahr jünger waren als die Stichprobe von Wallach und Kogan und da sie nicht einzeln und mündlich getestet wurden, sind die Ergebnisse wieder nicht vergleichbar.Abgesehen davon ist es auch fraglich, ob für vergleichbare nordamerikanische und bundesdeutsche Stichproben überhaupt gleiche Ergebnisse in einem solchen Test erwartet werden können.
Krause teilt lediglich mit, daß zwischen Kreativität und sozioökonomischem Status keine Beziehung besteht. Daß diese Angabe weder ausreichend informativ noch überzeugend ist, wurde bereits kritisiert (vgl.S. 1o6-1o7).

Während also aufgrund der in der Literatur enthaltenen Angaben nicht beurteilt werden kann, ob man für Kinder aus verschiedenen Sozialschichten mit unterschiedlichen Kreativitätstestergebnissen rechnen muß, liegen einige Hinweise darauf vor, daß die interne Konsistenz des Tests von Wallach und Kogan und seine relative Unabhängigkeit von Intelligenztestwerten auch für Stichproben mit niedrigerem sozioökonomischen Status nachgewiesen werden können (Ward, Kogan und Pankove 1972 an 191 überwiegend farbigen Fünftkläßlern, Massad 1969 an 93 Sechstkläßlern aus Mittelschicht- und 39 Sechstkläßlern aus Grundschichtfamilien).

Ein abweichendes Ergebnis stammt aus der Untersuchung von Kogan und Morgan (1969): In der spielähnlichen Untersuchungssituation korrelierten die Anzahlwerte der beiden Subtests von Wallach und

Kogan für Jungen und Mädchen aus Familien der unteren Mittelschicht nicht signifikant. Da die beiden Probandenzahlen sehr gering waren (25 Jungen und 31 Mädchen wurden in spielähnlichem Kontext getestet) und nur 2 der 5 Untertests verwendet wurden, ist dieses Resultat vermutlich weniger zuverlässig als die übereinstimmenden Ergebnisse der beiden anderen Untersuchungen.

Eine zusammenfassende Beurteilung der vereinzelten Ergebnisse zur Frage der Schichtabhängigkeit des Verfahrens von Wallach und Kogan kann somit ebenfalls nur feststellen, daß sie keine sichere Antwort ermöglichen und daß weitere Untersuchungen erforderlich sind.

2.335 Zusammenhänge zwischen den Ergebnissen des Verfahrens von Wallach und Kogan, Intelligenz und Schulleistung

Während Wallach und Kogan aufgrund ihrer Kreativitätskonzeption und ihrer Untersuchungsergebnisse von Nullkorrelationen zwischen Kreativitätstest- und Intelligenztestergebnissen ausgingen und die Frage nach eventuellen Zusammenhängen zwischen Kreativität und Schulleistungen gar nicht stellten,[1] wurden in einer Reihe späterer Untersuchungen niedrige bis mäßige Kreativitäts-Intelligenz-Korrelationen gefunden[2] (z.B. von Grote, Hajek u.a. 1969, Krause 1972, Richards 1974) und einige schwache, aber signifikante Zusammenhänge zwischen Kreativitätstestscores und Zensuren (Grote, Hajek u.a. 1969, Krause 1972, Murphy 1973). Wenn erstere zusätzlich zu Intelligenztestergebnissen als Prädiktoren von Schulleistungsmaßen verwendet werden, verbessert sich die Vorhersage nur unerheblich.

Ergebnisse von Untersuchungen über die Persönlichkeit Kreativer, die z.B. von Cattell und seinen Mitarbeitern auch auf Schüler ausgedehnt wurden,[3] machen es wahrscheinlich, daß kreative Kinder keine bequemen Schüler sind. Daraus wurde häufig die Hypothese

---

1) Wallach und Kogan untersuchten allerdings die Beziehungen zwischen Kreativität, Intelligenz und schulleistungsrelevanten Variablen wie Konzentration auf und Interesse am Unterrichtsstoff u.ä.; die Ergebnisse belegen die überragende Bedeutung der Intelligenz für diese Variablen.

2) Mögliche Gründe für diesen Widerspruch wurden bereits erörtert, s. dazu S. 95-96.

3) S. dazu z.B. Cattell und Butcher 1968

abgeleitet, daß Lehrer kreative Schüler nicht schätzen, und einige amerikanische Untersuchungsergebnisse scheinen das zu bestätigen (z.B. Bachtold 1974, Getzels und Jackson 1958, Holland 1959, Torrance 1964 b, letzterer zitiert nach Ulmann 1968).

Zwei mit deutschen Schülern durchgeführte Untersuchungen kommen zu einem abweichenden Ergebnis: Sowohl von Grote, Hajek u.a. (1969) als auch von Krause (1972) wurden keine negativen Korrelationen zwischen den Kreativitätstestergebnissen und den Lehrerratings für die "Bravheit" der Schüler gefunden. Bei Krause korrelierte der Kreativitätsgesamtwert .23 mit dem Lehrerurteil über die Bravheit der Schüler; genau denselben Koeffizienten erhielten Grote, Hajek u.a. für die Beziehung der Lehrerratings zu Kreativität und Bravheit ihrer Probanden (dieser Wert ist bei Krause signifikant mit einer Zufallswahrscheinlichkeit von einem Prozent, bei Grote, Hajek u.a. insignifikant).
Da auch die letztgenannten Autoren signifikante positive Korrelationen zwischen der Bravheitsschätzung des Lehrers und einigen Kreativitätssubtestergebnissen fanden, kann zumindest für deutsche Stichproben nicht von einer negativen Beziehung zwischen Kreativitätstestergebnissen und dem Lehrerurteil über Bravheit, schulische Angepaßtheit o.ä. ausgegangen werden.

2.34 Schlußfolgerungen für die weitere Arbeit

Der Kreativitätstest von Wallach und Kogan wurde auf der Grundlage einer assoziationstheoretischen Kreativitätskonzeption entwickelt, die eine Konstruktvalidierung ermöglicht..
Der von den Testautoren durchgeführte Teil der Konstruktvalidierung verlief zum großen Teil erfolgreich; die in ihrer eigenen Untersuchung bestätigte Hypothese einer völligen Unabhängigkeit der Kreativitäts- von Intelligenztestergebnissen wurde allerdings durch die Resultate anderer Arbeiten widerlegt.
Den bisher überwiegend positiven Resultaten der Konstruktvalidierung stehen vergleichsweise wenig Untersuchungsergebnisse gegenüber, die die Übereinstimmungs- oder Vorhersagegültigkeit belegen; weitere Untersuchungen zu den verschiedenen Validitätsaspekten erscheinen dringend erforderlich.

Im Hinblick auf die Zielsetzung der vorliegenden Arbeit stellt sich die Frage, ob der Kreativitätstest von Wallach und Kogan für den Gebrauch mit deutschen Schülern adaptiert werden soll.
Bei dem Torrance-Test war die Beantwortung derselben Frage wegen seiner fehlenden theoretischen Fundierung und seiner bis jetzt unbefriedigenden Validierung leichtgefallen; der Verzicht auf eine deutsche Adaptation des Tests von Wallach und Kogan bedarf dagegen einer etwas ausführlicheren Begründung:
Das Verfahren ist mit 39 Items, die alle mit einer möglichst großen Zahl von Lösungen beantwortet werden sollen, zu lang. Vor allem bei einer Untersuchung ohne oder mit sehr großzügiger Zeitbegrenzung dürfte die Produktivität der Probanden stark von motivationalen Variablen beeinflußt werden. Wenn es sich dabei nur um aufgabenbezogene intrinsische Motivationen handeln würde, könnte dies sogar eine Erhöhung der Übereinstimmungs- und/oder Vorhersagevalidität bewirken. Das ist jedoch wenig wahrscheinlich, vielmehr muß die Möglichkeit gesehen werden, daß die allgemeine schulische Anstrengungsbereitschaft auch die Ergebnisse solcher Tests beeinflußt. (Die positive Korrelation zwischen Kreativitätstestergebnis und dem Lehrerurteil über die Bravheit der Schüler ist eventuell auf diese Weise zu erklären.)
Hinzu kommt, daß die Anforderungen, die der Test von Wallach und Kogan an die Kinder stellt, ziemlich gleichförmig sind. Alle Aufgaben sind verbal zu beantworten. Vier der fünf Subtests bestehen aus 8 bis 1o formal gleichen Items; die beiden Aufgabengruppen mit visuellen Aufgabenstimuli - es sind insgesamt 17 - verlangen von den Kindern genau dieselbe Aktivität, nämlich die Deutung von abstrakten Mustern bzw. Linien. Dies führt mit großer Wahrscheinlichkeit bei Gruppenuntersuchungen, möglicherweise aber auch bei Invididualtestungen, dazu, daß viele Kinder bei den letzten Items eines Subtests und eventuell auch bei den Subtests, die gegen Ende der Untersuchung bearbeitet werden, wesentlich weniger Antworten geben, als ihnen unter günstigeren Testbedingungen einfallen würden. Nur Schüler mit außergewöhnlicher Persistenz und die vermutlich nicht gerade zahlreichen Kinder, die durch alle 39 Aufgaben von Wallach und Kogan intrinsisch motiviert werden, dürften hier auszunehmen sein.
Somit besteht bei dem Test von Wallach und Kogan die Gefahr, daß nicht nur kognitive Voraussetzungen für kreatives Verhalten, sondern

auch extrinsische Motivationen erfaßt werden, wodurch die Übereinstimmungs- und Vorhersagevalidität des Verfahrens beeinträchtigt wird.

Obwohl aus diesem Grund darauf verzichtet wird, den Test von Wallach und Kogan für deutsche Schüler zu adaptieren, sollen bei der Entwicklung und Validierung eines Verfahrens für deutsche Schüler eine Reihe von mehr oder weniger gesicherten Schlußfolgerungen berücksichtigt werden, die aus den Ergebnissen der mit dem Wallach-Kogan-Test durchgeführten Untersuchungen abgeleitet wurden.

Die wichtigsten dieser Folgerungen sind:

- Die wesentlichen Punkte der assoziativen Kreativitätstheorie von Mednick und Wallach und Kogan bilden eine tragfähige Grundlage für die Konstruktion eines Verfahrens zur Erfassung kognitiver Voraussetzungen für kreatives Verhalten.
- Zwischen den Ergebnissen eines solchen Verfahrens und Intelligenztestwerten sind mäßige positive Korrelationen zu erwarten. Der Nachweis, daß es sich um zwei verschiedene Persönlichkeitsdimensionen handelt, muß auf andere Weise geführt werden als mit Hilfe von Nullkorrelationen.
- Individualtestung in spielähnlichem Kontext, ohne Zeitbegrenzung und mit mündlicher Aufgabenbeantwortung scheint keine unabdingbare Voraussetzung für die Durchführung eines Kreativitätstests zu sein. Gruppentestungen in entspannter, nicht schulähnlicher Atmosphäre und mit großzügig bemessener Testzeit können nach den vorliegenden Untersuchungsergebnissen ebenfalls zu validen Meßwerten führen.
- Es kann nicht davon ausgegangen werden, daß Jungen und Mädchen in gleicher Weise auf einen Kreativitätstest reagieren. Neben geschlechtsspezifischen Unterschieden in den Testergebnissen ist eventuell auch mit differierenden Beziehungen dieser Werte zu anderen Persönlichkeitsvariablen zu rechnen.
- Auch die Möglichkeit schichtabhängiger Unterschiede in den Testergebnissen muß überprüft werden.
- Schulleistungen sind keine geeigneten Validitätskriterien für einen Kreativitätstest.

Empirischer Teil

3 Entwicklung und Validierung eines Verfahrens zur Erfassung divergenten Denkens für Schüler der BRD

Vor der Darstellung der Entwicklung und Validitätsprüfung des Verfahrens wird seine theoretische Fundierung besprochen.

3.1 Theoretische Fundierung

Die Entwicklung des Tests zur Erfassung des divergenten Denkens basiert auf folgenden theoretischen Annahmen:

a) Voraussetzung für kreatives Verhalten und kreative Leistungen[1] ist das Zusammenwirken von kognitiven Fähigkeiten, motivationalen Faktoren und Persönlichkeitsmerkmalen im engeren Sinn.

Selbst eine reliable und valide Messung einer dieser Faktorengruppen allein ermöglicht keine hinreichend sichere Diagnose der "Kreativität" und keine zuverlässige Prognose kreativer Leistungen.

b) Zu den für kreative Leistungen bedeutsamen kognitiven Faktoren gehören in erster Linie die Fähigkeiten des divergenten Denkens. Unter divergentem Denken werden in Anlehnung an Guilford diejenigen Denkprozesse verstanden, die zu einer Vielzahl von Lösungsmöglichkeiten für ein bestimmtes Problem führen - im Gegensatz zum konvergenten Denken, das zu einer einzigen richtigen Lösung führt. Je nach Aufgaben- oder Problemstellung führt die eine oder andere Denkform zum Ziel; für komplexere Problemlösungsprozesse sind allerdings meist divergentes und konvergentes Denken erforderlich.

Guilford unterscheidet 4 "Inhaltsbereiche", in denen Denkprozesse ablaufen können.[2] Ob alle Denktätigkeiten einer dieser Inhaltsklassen eindeutig zugeordnet werden können und ob die 4 Kategorien ein vergleichbares Abstraktionsniveau besitzen, soll hier nicht erörtert werden, wichtig erscheint dagegen die starke Berücksichtigung nichtverbaler Intelligenzfaktoren in Guilfords Modell. Ein Meßinstrument, das Denkprozesse in 4 verschiedenen In-

---

1) Auf die Problematik der Kreativitätsdefinition selbst soll hier nicht eingegangen werden (s. dazu die ausführlichen Diskussionen von Krause 1972, Seiffge-Krenke 1974, Ulmann 1968); den meisten Definitionen ist die Aussage gemeinsam, daß kreatives Verhalten zu neuen Ergebnissen führt, die in irgendeiner Weise adäquat, nützlich oder überzeugend sein müssen (s. dazu auch Lohmann 1975 und Mainberger 1975).

2) Vgl. dazu S. 4

haltsbereichen verlangt, erhebt wahrscheinlich eine repräsentativere Stichprobe intellektuellen Verhaltens als ein ausschließlich verbal zu beantwortendes Verfahren. Außerdem dürfte es eine bessere intrinsische Motivierung derjenigen Probanden bewirken, deren Begabungs- und/oder Interessenschwerpunkt nicht im verbalen Bereich liegt. Guilfords Klassifizierung der Denkprozesse nach dem Kriterium der dabei entstehenden "Produkte" erscheint dagegen weniger zwingend.[1] Außerdem dürfte die Bedeutung dieser Klassifikation für die Denkleistungen von Kindern nicht in allen Punkten nachgewiesen sein.

c) Guilfords Modell der Struktur des Intellekts beinhaltet keine Theorie des divergenten Denkens, d.h. er stellt keine Hypothesen darüber auf, wie es zu einer Vielzahl von Antworten oder zu einer neuen, besseren Lösung eines bekannten und bereits gelösten Problems kommt. Ansätze zur Beantwortung dieser Fragen bietet dagegen die assoziationstheoretische Kreativitätskonzeption von Mednick und Wallach und Kogan. Divergentes Denken wäre nach dieser Theorie aufzufassen als die Herstellung von Beziehungen zwischen vorher unverbundenen Bestandteilen des Assoziationsspeichers. Je häufiger solche Beziehungen gestiftet werden und je größer die assoziative Distanz zwischen den neu verbundenen Einheiten vorher war, desto höher ist die jeweilige Fähigkeit zum divergenten Denken einzuschätzen.[2]

d) Im Gegensatz zu der Auffassung von Wallach und Kogan wird eine mäßige positive Korrelation zwischen dieser assoziationstheoretisch konzipierten Fähigkeit des divergenten Denkens und den Ergebnissen traditioneller Intelligenztests erwartet.
Diese Annahme gründet sich auf folgende Überlegungen:

Intelligenztestergebnisse hängen u.a. direkt vom Inhalt des Assoziationsreservoirs ab.
Nach Mednick (1962) und Wallach und Kogan assoziieren Hochkreative insgesamt mehr Ideen als wenig Kreative, woraus auf ihr größeres Assoziationsreservoir geschlossen werden könnte. Krause (1972), der diese Schlußfolgerung übernimmt, erkennt die aus dieser Position resultierende Schwierigkeit, die beiden Konstrukte Kreativität und Intelligenz gegeneinander abzugrenzen, und führt

---

1) Vgl. dazu S. 6
2) S. dazu S. 78

die erforderliche Klärung durch Einbeziehung einer zweiten Variablen herbei: "Hochkreative unterscheiden sich von Niedrigkreativen nicht nur in der Anzahl der kognitiven Elemente, die sie zu einem bestimmten Problem erworben haben. Eine weitere zu berücksichtigende Variable ist die Verfügbarkeit, respektive Abrufbarkeit dieser kognitiven Elemente. Sie können wohl gespeichert, aber nicht zugänglich sein." (S.74)

Die Frage, wie sich die Assoziationsreservoirs hochkreativer Individuen von denjenigen hochintelligenter, aber wenig kreativer Personen unterscheiden, wird nicht erörtert. Nach Krauses Aussagen (s.o.) ist anzunehmen, daß er den Hochkreativen nicht nur leichter zugängliche, sondern auch eine größere Zahl von Assoziationen zugesteht als den wenig kreativen, aber hochintelligenten Individuen.

Wegen der positiven Korrelation zwischen Intelligenz- und Kreativitätstestergebnissen erscheint es jedoch plausibler, nur die Verfügbarkeit bzw. Abrufbarkeit der Einfälle als die zwischen diesen beiden Gruppen differenzierende Variable zu betrachten; bei vergleichbarem Umfang der Assoziationsreservoirs würden die Kreativen infolge ihrer weniger starken Abwehrmechanismen bessere Ergebnisse in Kreativitätstests erzielen als intelligente, aber wenig kreative Personen.[1] Die Zahl der insgesamt vorhandenen assoziativen Einheiten wäre aber sowohl für Intelligenz- als auch für Kreativitätstests eine wichtige Bestimmungsgröße, da nur abgerufen werden kann, was im Speicher vorhanden ist. Diese Variable wäre somit verantwortlich für die positive Korrelation zwischen Intelligenz- und Kreativitätstestergebnissen.

e) Aus der Hypothese, daß sich die Stärke der Abwehrmechanismen auf die kognitiven Voraussetzungen kreativen Verhaltens auswirkt, ergibt sich die Annahme einer relativ engen Beziehung zwischen Kreativitätstestergebnissen und verschiedenen Persönlichkeitsvariablen, die ihrerseits mit den Abwehrmechanismen zusammenhängen.

Aus den dargestellten theoretischen Annahmen ergeben sich für die Testkonstruktion folgende Konsequenzen:

---

1) Intelligenztestergebnisse sind gegen Abwehrmechanismen deshalb nicht oder kaum anfällig, weil die Antwortalternativen meist vorgegeben werden, so daß die richtige Lösung nur erkannt und nicht erinnert werden muß.

Zu a) Das Verfahren soll wesentliche kognitive Voraussetzungen für kreatives Verhalten, nicht jedoch "Kreativität" erfassen. Deshalb wird es - im Gegensatz zur verbreiteten Terminologie - als Test zur Erfassung des divergenten Denkens und nicht als Kreativitätstest bezeichnet.[1)]
Daß es nur eine von mehreren Voraussetzungen kreativer Leistungen messen soll, erschwert die Validierung, weil als Ergebnis von Überprüfungen der Übereinstimmungsvalidität nicht mit Koeffizienten gerechnet werden kann, die als sichere Validitätsbelege interpretiert werden können. Schon aus diesem Grund erscheint es notwendig, mehrere verschiedenartige Validierungsuntersuchungen durchzuführen.

Zu b) Da das Verfahren Fähigkeiten des divergenten Denkens messen soll, müssen die Aufgaben aus offenen Fragen bestehen, für die eine möglichst große Zahl von Antworten gefordert wird. Die Schwierigkeiten, die durch die freie Aufgabenbeantwortung für die Auswertung und die Standardisierung des Tests entstehen, müssen in Kauf genommen werden.
Um die Repräsentativität der Verhaltensstichprobe zu erhöhen, soll der Test nicht nur verbale Aufgaben enthalten, sondern nach Möglichkeit Items aus den 4 Inhaltsbereichen Guilfords.
Da alle Aufgaben mit einer möglichst großen Zahl von Lösungen beantwortet werden sollen, erscheint es auch im Hinblick auf die für ein solches Verfahren besonders wichtige intrinsische Motivierung günstig, einen Test mit relativ verschiedenartigen und damit abwechslungsreichen Items zu konstruieren.

Zu c) Die assoziationstheoretische Kreativitätskonzeption von Mednick und Wallach und Kogan scheint eine tragfähige Grundlage für die Konstruktion des Verfahrens zu bieten. Wie Guilfords Konzept der divergenten Produktion führt sie zu Items mit freier Aufgabenbeantwortung, aber - im Gegensatz zu den DP-Tests von Guilford - zu Aufgaben mit assoziativem Charakter, d.h. die Fragen erlegen den Probanden nur minimale Beschränkungen auf, so daß die meisten von ihnen mit mehreren Lösungen antworten können.
Die assoziationstheoretische Kreativitätskonzeption impliziert ferner - ebenfalls im Gegensatz zu Guilford -, daß ein Kreativitätstest ohne Zeitbegrenzung durchgeführt werden müsse. Da die Unter-

---

1) S. dazu auch Krause 1972 (S. 84) und Nicholls 1972

suchungsergebnisse die zu diesem Punkt aufgestellten Hypothesen nicht eindeutig stützen[1] und nachdem das neu zu entwickelnde Verfahren nur Fähigkeiten des divergenten Denkens - möglichst wenig konfundiert durch aufgabenfremde Motivationen - erfassen soll, wird diese Implikation folgendermaßen modifiziert: Für die Bearbeitung der Aufgaben soll so viel Zeit zur Verfügung stehen, daß die Schreibgeschwindigkeit der Kinder ihre Produktivität nicht beeinflußt. Außerdem soll die Einhaltung der Zeitgrenzen möglichst unmerklich realisiert werden.

Zu d) Für das zu konstruierende Verfahren wird keine Nullkorrelation, sondern ein mäßiger positiver Zusammenhang mit Intelligenztestergebnissen erwartet. Die erforderliche Abgrenzung gegen die traditionellen Intelligenztests soll durch den Nachweis erfolgen, daß die z.T. sehr unterschiedlichen Einzelitems untereinander stärker korrelieren als mit den Subtestergebnissen von Intelligenztests. Dieser Nachweis kann mit Hilfe der Faktorenanalyse und durch einen Vergleich des Mittelwertes der Interkorrelationen der Einzelitems mit der durchschnittlichen Korrelation zwischen diesen Items und den Intelligenzsubtests erbracht werden.
Die Forderung, daß die durchschnittliche Interkorrelation der Einzelitems höher sein muß als die durchschnittliche Korrelation der Items mit den Intelligenzsubtests, stellt wegen der niedrigen Reliabilität von Einzelitems ein überaus strenges Kriterium dar. Es wird allerdings angenommen, daß ein Item, das mit einer größeren Zahl von Lösungen beantwortet werden kann, eine höhere Reliabilität besitzt als eine richtig oder falsch zu beantwortende Einzelaufgabe, aber die Reliabilität liegt vermutlich unter derjenigen eines aus 1o oder mehr Richtig-Falsch-Items bestehenden Subtests. Falls die Überprüfung ein positives Resultat erbringt, kann sie als überzeugender Beleg für die Konstruktvalidität des neuen Verfahrens gewertet werden. Ein negatives Resultat würde schwerwiegende Probleme für die Validierung aufwerfen und u.U. dazu führen, daß die Forderung einer vergleichsweise kurzen Testdauer aufgegeben werden muß.

Zu e) Eine Untersuchung des Zusammenhangs zwischen Maßen des diver-

---

1) S. dazu S.118-129, vor allem S. 125-126

genten Denkens und Fragebogenscores für die Tendenz zur Verdrängung erscheint wenig erfolgversprechend; nach den Ergebnissen Krauses[1] kann nicht damit gerechnet werden, beim derzeitigen Forschungsstand zu Resultaten zu gelangen, die über die untersuchten Stichproben hinaus generalisiert werden können.

Dagegen kann die Hypothese aufgestellt und überprüft werden, daß zwischen den Ergebnissen des zu entwickelnden Verfahrens und Meßwerten für Persönlichkeitsmerkmale, die weniger stark durch den Unterrichtsstil des Klassenlehrers beeinflußt werden, engere Beziehungen bestehen als zwischen Intelligenztestergebnissen und diesen Persönlichkeitsvariablen. Vor allem aber könnten Hypothesen über Richtung und Enge des Zusammenhangs zwischen verschiedenen kreativitätsrelevanten Persönlichkeitsmerkmalen und den Ergebnissen der Versuchsform VI, K aufgestellt und überprüft werden. Die Bestätigung dieser Untersuchungshypothesen könnte als weiterer Hinweis auf die Konstruktvalidität des neuen Tests zur Erfassung des divergenten Denkens betrachtet werden.

Die dargestellte theoretische Konzeption des zu entwickelnden Tests ermöglicht trotz ihres vorläufigen und unvollständigen Charakters erste Ansätze zu einer Konstruktvalidierung, die angesichts der zu erwartenden wenig überzeugenden Übereinstimmungsvalidität[2] und vor allem wegen der Fragwürdigkeit der für Schulkinder zur Verfügung stehenden Außenkriterien kreativen Verhaltens unbedingt erforderlich erscheint.

3.2 Testentwicklung

Das Verfahren wurde in der Zeit vom Herbst 1972 bis Ende 1975 im Zentrum für empirische pädagogische Forschung der Erziehungswissenschaftlichen Hochschule Rheinland-Pfalz entwickelt. Tabelle 6 gibt einen Überblick über die im Rahmen der Testentwicklung durchgeführten Untersuchungen.

---

1) S. dazu S. 107-108

2) S. dazu die Erörterung zu Punkt a), S. 147

Tabelle 6: Überblick über den Untersuchungsablauf

Untersuchungszeitpunkt	Kurzbezeichnung der Versuchsform*)	Stichprobe bzw. Pbn	Untersuchungsziele
Herbst 1972	ohne Bezeichnung 68 Aufgaben	2 Schüler aus 4. und 2 Schüler aus 5. Klassen	Eliminierung ungeeigneter Aufgaben
Herbst 1972	Versuchsform II, 4 (31 Aufgaben)	312 Schüler 4. Klassen	Feststellung trennscharfer und relativ geschlechts- und schichtunabhängiger Aufgaben
	Versuchsform II, 5 (27 Aufgaben)	264 Schüler 4. Klassen	
Herbst 1973	Versuchsform IV, 1 (11 Aufgaben)	526 Schüler 5. Klassen aus Gesamtschulen	Überprüfung von Trennschärfe, interner Konsistenz und diskriminanter Validität, Untersuchung der Korrelation mit Schulleistungen
Frühjahr 1974		130 Schüler 5. Hauptschulklassen	Überprüfung der Übereinstimmungsvalidität

*) Die Untersuchungen wurden z.T. an denselben Stichproben wie die Erprobung der Vorformen des Bildungs-Beratungs-Tests von Ingenkamp, Wolf u.a. durchgeführt; die Benennung der Versuchsformen richtete sich nach derjenigen der BBT-Vorformen

Wie aus der Übersicht zu ersehen ist, wurden bereits während der Entwicklungsphase erste Überprüfungen der Validität des Verfahrens vorgenommen, wenngleich der größte Teil der Validitätsprüfungen mit der Endform VI, K durchgeführt wurde.

3.21 Erste und zweite Versuchsform (II, 4 und II, 5)

Im Herbst 1972 wurden die ersten beiden Testvorformen zusammengestellt. Sie bestanden aus Aufgaben, die verschiedene Formen des divergenten Denkens erfassen sollten und dementsprechend eine freie Aufgabenbeantwortung erforderten.
Entsprechend der assoziationstheoretischen Kreativitätskonzeption waren die Items so beschaffen, daß die Mehrzahl der Kinder mit einer

größeren Zahl von Lösungen antworten kann. Es wurden Aufgaben neu konstruiert, andere wurden aus amerikanischen Kreativitätstests von Guilford, Torrance und Wallach und Kogan - meist in modifizierter Form - übernommen.[1)]

68 Einzelaufgaben wurden an einigen Viert- und Fünftkläßlern in face-to-face-Testungen erprobt, wobei mehrere ungeeignete, meist zu schwierige Aufgaben eliminiert oder abgeändert wurden. Die restlichen 58 Aufgaben wurden auf 2 Versuchsformen - II, 4 und II, 5 - verteilt.

Um Verfälschungen der Testergebnisse durch Ermüdung, die bei Testzeiten von mehr als 7o Minuten zu erwarten waren, möglichst gering zu halten, wurden die Items jeder Versuchsform in 3 etwa gleich lange Teile A, B und C aufgeteilt und so angeordnet, daß etwa ein Drittel der Versuchspersonen Teil A an erster, das zweite Drittel A an zweiter und das letzte Drittel A an dritter Stelle erhielt; analog wurde mit den beiden anderen Testteilen verfahren.

Beide Versuchsformen enthielten Aufgaben aus den 4 Inhaltsbereichen Guilfords.[2)]

Versuchsform II, 4, bestehend aus 31 Aufgaben, wurde an 312 Schülern aus 9 vierten Klassen rheinland-pfälzischer Grundschulen erprobt; Versuchsform II, 5 enthielt 27 Aufgaben und wurde in 8 vierten Klassen mit 264 Schülern aus Rheinland-Pfalz durchgeführt.

Als Testleiter fungierten Lehrerstudenten, die über die speziellen Erfordernisse dieser Testsituation informiert worden waren. Die beiden Vorformen wurden nach Produktivität, d.h. nach der Anzahl der akzeptablen Lösungen, und nach Originalität - operationalisiert als statistische Seltenheit der jeweiligen Antwort - ausgewertet. Die Auswerter erhielten vorläufige Auswertungsrichtlinien, die u.a. die Akzeptierung origineller, aber aus der Sicht Erwachsener logisch oder sachlich nicht ganz richtiger Antworten gewährleisten sollten. Diese Richtlinien mußten infolge der großen Zahl unvorhergesehener Lösungen mehrmals modifiziert und vor allem ergänzt werden.

---

1) Ein wesentlicher Teil der Neukonstruktionen und Modifikationen geht auf Vorschläge von Herrn Dipl.-Päd. Hans Haenisch zurück.

2) S. Anhang A 1

Originalitätspunkte erhielten bei den Versuchsformen II, 4 und II, 5 alle Antworten, die in der jeweiligen Stichprobe von höchstens 5 Prozent der Probanden gegeben wurden. Diese sehr zeitaufwendige Prozedur erforderte für jede verbale Aufgabe eine alphabetische Auflistung aller von den Schülern gegebenen Antworten und die Feststellung der Häufigkeit, mit der jede Antwort gegeben wurde. Dabei wurde nicht - wie bei Krause 1972 - jede Antwort einzeln aufgelistet,[1] sondern inhaltlich übereinstimmende Antworten wurden zu Kategorien zusammengefaßt; dadurch wurde verhindert, daß häufig geäußerte Antworten dann einen Originalitätspunkt bekommen, wenn sie in einer für diese Altersstufe seltenen Formulierung gegeben werden.

An den Stichproben, die die Versuchsformen II, 4 und II, 5 bearbeitet hatten, waren außerdem folgende Daten erhoben worden:

Geschlecht,
Altersstufe (9 Stufen),
Zahl der Geschwister,
Beruf des Vaters und der Mutter (Schichteinteilung nach Kleining und Moore 1968),
die Zensuren in Deutsch
Mathematik und
Zeichnen,
eine Schätzung der Intelligenz durch den Lehrer, außerdem Ratings der Phantasie, des "schöpferischen Denkens", der Häufigkeit origineller Einfälle, der sozialen Anpassung, Disziplin und Schulangst.

Aufgrund der unbefriedigenden Untersuchungsergebnisse zur Validität von Lehrerurteilen über Fähigkeiten des divergenten Denkens[2] wurden die Ergebnisse der Lehrerratings nur sehr selten und immer mit sehr schwacher Gewichtung als Entscheidungshilfe für die Beibehaltung bzw. Eliminierung von Aufgaben der beiden Versuchsformen herangezogen.

Nachdem die meisten Aufgaben in amerikanischen Kreativitätstests - wenn auch viele in leicht modifizierter Form - erprobt und Hinweise für ihre Validität gefunden worden waren, war die Trennschärfe[3]

---

1) Vgl. S. 1o6-1o7

2) S. dazu S. 4o-41

3) Da die Aufgaben nicht richtig oder falsch zu lösen sind, wurden statt der punktbiserialen Korrelation Produkt-Moment-Korrelationskoeffizienten zwischen der jeweiligen Anzahl akzeptabler Antworten und der Gesamtpunktzahl berechnet. Der niedrigste Koeffizient der akzeptierten Aufgaben betrug .45, der Rest lag über .5o, der höchste Wert betrug .61. Zur Berechnung von Produkt-Moment-Korrelationen s. Anmerkung 1, S. 155

der Items das wichtigste Kriterium für ihre Eliminierung oder weitere Verwendung.
Außerdem wurden Aufgaben ausgeschieden, die nicht geschlechts- oder schichtneutral waren.

Die Items zur Erfassung des symbolischen Denkens und diejenigen aus dem Verhaltensbereich hatten in der Mehrzahl sehr niedrige Trennschärfen, was eventuell darauf zurückzuführen sein könnte, daß sie im Verhältnis zu den semantischen und figuralen Aufgaben stark unterrepräsentiert waren. Da aber sowohl die symbolischen als auch die Verhaltens-Items auch untereinander meist keinen signifikanten Zusammenhang zeigten, wurde die Zielsetzung, die Fähigkeit zu divergentem Denken in den 4 Inhaltsklassen Guilfords zu erfassen, aufgegeben. Ziel der weiteren Entwicklungsarbeit war die Konstruktion eines Meßverfahrens der Fähigkeit zu divergentem Denken für Neun- bis Zwölfjährige, das nur den semantischen und den figuralen Inhaltsbereich berücksichtigt.

3.22 Dritte Versuchsform (IV, 1)

Die nächste Versuchsform (IV, 1) wurde dementsprechend aus verbalen und figuralen Aufgaben der Versuchsformen II, 4 und II, 5, die eine Trennschärfe von mindestens .45 hatten,[1] zusammengestellt. Da kreatives Verhalten an unseren Schulen bestenfalls eine untergeordnete Rolle spielt und die Lehrer demzufolge wenig geneigt sind, Unterrichtszeit für die Erprobung eines Verfahrens zur Verfügung zu stellen, an dessen Ergebnissen im Rahmen des Unterrichts wenig Interesse besteht, mußte die neue Versuchsform möglichst kurz sein, d.h. sie durfte insgesamt höchstens eine Unterrichtsstunde beanspruchen.

Die Versuchsform IV, 1 enthält daher nur 11 Aufgaben.[2]

Die erste der 11 Aufgaben, die zeichnerische Lösungen verlangt, wurde nur nach Produktivität, d.h. nur nach der Anzahl der akzeptablen Lösungen ausgewertet.
Die Antworten der Schüler zu den übrigen 1o Testaufgaben wurden wieder nach Anzahl der akzeptablen Lösungen und nach Originalität ausgewertet.

---

1) Vgl. Anmerkung 3, S. 152
2) S. Anhang A 2

Die Originalitätsauswertung orientierte sich zwar immer noch an der statistischen Seltenheit der Antworten, aber das bei den Versuchsformen II, 4 und II, 5 angewandte Verfahren wurde nach einem Vorschlag von Vernon (1971) folgendermaßen abgeändert:

Aus der Stichprobe von 444 Schülern wurden mit Hilfe von Zufallszahlen aus allen Klassen so viele Schüler entnommen, bis eine Zufallsstichprobe von 1oo Probanden vorlag. Für jede der 1o Aufgaben wurden die akzeptierten Antworten dieser 1oo Schüler alphabetisch aufgeführt und die Häufigkeit, mit der jede Antwort vertreten war, festgehalten. Antworten, die in dieser Zufallsstichprobe 1-5 Mal gegeben wurden, erhielten 2 Originalitätspunkte, Antworten, die 6-14 Mal vorkamen, einen, und alle Antworten, die häufiger als 14 Mal auftraten, keinen Originalitätspunkt.[1)]
Durch Verwendung der von Vernon vorgeschlagenen Originalitätsauswertung wurde erreicht, daß sich die Mittelwerte und Varianzen der Anzahl- und Originalitätswerte bei der Versuchsform IV, 1 nicht so stark unterschieden wie bei den Aufgaben der Versuchsform II, 4 und II, 5, bei denen nur Antworten, die von höchstens 5 Prozent der Kinder gegeben worden waren, einen Originalitätspunkt erhielten.

3.221 Untersuchung an 15 Gesamtschulklassen

Die Versuchsform IV, 1 wurde im Herbst 1973 in 15 Gesamtschulklassen an 526 Fünftkläßlern durchgeführt. 444 von diesen Schülern bearbeiteten außerdem einen Schulleistungstest (Allgemeiner Schulleistungstest AST 4[2)]) einen Intelligenztest (Primary Mental Abilities PMA in der Übersetzung von Kemmler und Langheinrich 1967) und 18 Untertests des zu erprobenden Bildungs-Beratungs-Tests BBT, eines im Zentrum für empirische pädagogische Forschung entwickelten Intelligenztests von Ingenkamp, Wolf u.a.. Von den übrigen Daten, die an dieser Stichprobe erhoben worden waren, wurden nur die Noten in Deutsch, Mathematik und der ersten Fremdsprache nach dem ersten Halbjahr der 5. Klasse zu den Ergebnissen der Versuchsform IV, 1 in Beziehung gesetzt.

Die Aufgaben der Versuchsform IV, 1 wurden im Hinblick auf folgende Fragen überprüft:

---

1) Zur Begründung s. S.129-132
2) Fippinger 1967

1. Übersteigen die Interkorrelationen der Testitems untereinander die Korrelationen der Items mit den Intelligenztestaufgaben? Diese Frage nach der diskriminanten Validität der Versuchsform IV, 1 mußte auf jeden Fall positiv zu beantworten sein, wenn das Verfahren etwas anderes mißt als die beiden Intelligenztests.

2. Sind darüber hinaus Trennschärfe und innere Konsistenz der Aufgaben der Versuchsform IV, 1 so hoch, daß die Annahme, sie würden das gleiche Konstrukt messen, begründet erscheint? Auch auf diese Frage wurde eine positive Antwort erwartet.

3. Kann durch die Hinzunahme der Aufgaben von Versuchsform IV, 1 zu den Items des BBT die Vorhersage der Schulleistungen - gemessen durch Schulleistungstests und durch die Benotung - signifikant und substantiell verbessert werden? Aufgrund der in der Literatur berichteten Untersuchungsergebnisse wurde erwartet, daß das nicht der Fall sein würde.

Zu 1:

Zur Prüfung der ersten Frage wurden zunächst die 147 Produkt-Moment-Korrelationen[1] der PMA-Untertests mit den 21 Ergebnissen der Versuchsform IV, 1 (11 Produktivitäts- und 1o Originalitätsmaße) nach Fisher z-transformiert und gemittelt, ebenso die Korrelationen der 5 Untertests des BBT, die in die Endform aufgenommen wurden, mit den Aufgaben der Versuchsform IV, 1 und die 2oo Interkorrelationen der Items von IV, 1 untereinander.[2]

Die durchschnittlichen Korrelationskoeffizienten zeigt Tabelle 7.

---

1) Wie in den meisten amerikanischen und deutschen Untersuchungen mit Kreativitätstestergebnissen wurden auch die Testrohwerte dieser Arbeit wie Intervalldaten behandelt, obwohl sie wahrscheinlich nur Ordinalskalenniveau besitzen, weil für letztere z.T. keine geeigneten Programme zur elektronischen Datenverarbeitung vorliegen.(So verarbeitet z.B. das im Regionalen Hochschulrechenzentrum Kaiserslautern verfügbare Rangkorrelationsprogramm keine Probanden mit missing data, was bei großen Variablenzahlen meist eine erhebliche Reduktion des Stichprobenumfangs zur Folge hat.)

2) Da Produktivitäts- und Originalitätsmaße der gleichen Aufgaben sehr hoch miteinander korrelieren, wurden diese 1o Koeffizienten nicht in die Mittelwertsberechnung einbezogen.

Tabelle 7: Mittelwerte der Korrelationen der Ergebnisse der Versuchsform IV, 1 untereinander und mit den Untertests der PMA und des BBT. (N = 444)

	VI, 1	PMA	BBT
IV, 1	$\bar{r}$ = .300	$\bar{r}$ = .105	$\bar{r}$ = .139

Die Items der Versuchtsform IV, 1 korrelierten untereinander höher als mit den Untertests von PMA und BBT, der Unterschied ist allerdings statistisch nicht abgesichert. (Eine Signifikanzprüfung für die Differenz der Mittelwerte von Korrelationskoeffizienten scheint noch nicht vorzuliegen. Ein von Edwards 1971 beschriebenes Verfahren von Hotelling 1940 zur Signifikanzprüfung des Unterschieds zwischen zwei Korrelationskoeffizienten ist auf den vorliegenden Fall nicht anwendbar.)
Eine Überprüfung der einzelnen Koeffizienten der Korrelationsmatrix legt zwar ebenfalls eine positive Beantwortung der ersten Frage nahe, aber auch hier sind keine Angaben über die Wahrscheinlichkeit möglich, mit der diese Ergebnisse nach dem Zufall zu erwarten wären. Sämtliche Aufgaben der Versuchsform IV, 1 korrelieren untereinander signifikant auf dem 1 %-Niveau, von den 147 Korrelationskoeffizienten der Untertests der PMA mit den Items der Versuchsform IV, 1 sind 51 hochsignifikant, von den 105 Koeffizienten, die für den Zusammenhang zwischen BBT und Versuchsform IV, 1 berechnet wurden, erreichen 57 Signifikanz auf dem 1 %-Niveau.

Eine weitere Möglichkeit zur Überprüfung der ersten Frage bietet die Faktorenanalyse. Daher wurde eine Hauptachsen-Faktorenanalyse mit anschließender Varimax-Rotation für folgende Variablen gerechnet: Zensuren (Deutsch, Rechnen, Sachkunde), AST 4-Ergebnisse, 7 PMA-Untertests, 18 BBT-Untertests und 21 Testwerte der Versuchsform IV, 1[1]. Die Ergebnisse sind auszugsweise in Tabelle 8 wiedergegeben.

1) Programm PAFA des Deutschen Rechenzentrums; Autoren: P. Schnell und Dr. F. Gebhardt, DRZ Darmstadt. Das Programm verwendet als erste Kommunalitätenschätzungen die Quadrate der multiplen Korrelationen; die Rotation der Faktoren erfolgt nach dem verbesserten Varimax-Kriterium von Kaiser (1959).

Tabelle 8: Rotierte Faktorenmatrix der Schulleistungs- und Intelligenztestvariablen und der Ergebnisse der Versuchsform IV, 1. Bei letzteren wurden alle Ladungszahlen aufgeführt, bei den anderen Variablen nur diejenigen ab .3o. (N = 444)*

	1. Faktor	2. Faktor	3. Faktor	4. Faktor	h^2
Deutschnote			.67	-.35	.57
Mathematiknote	.38		.49	-.46	.6o
Sachkundenote			.57		.44
AST 4 Leseverständnis			-.64		.48
" Wortschatz			-.72		.6o
" Kopfrechnen	-.42		-.38	.5o	.58
" schriftl. Rechnen			-.3o	.65	.56
" Textaufgaben	-.43		-.43	.5o	.64
" Rechtschreibung			-.61	.42	.57
" Heimatkunde			-.61		.49
PMA Bilderwortschatz			-.59		.37
" Wortschatz			-.8o		.68
" Raumerfassung	-.56				.34
" Wortgruppen	-.53				.38
" Figurengruppen	-.66				.45
" Wahrnehmungstempo	-.37				.25
" Zahlentest				.63	.44
BBT Gegensätze	-.35		-.51		.4o
" Zahlenreihen	-.56		-.3o		.49
" Zahlenschrift	-.32		-.32	.42	.39
" Bildanalogie	-.6o				.42
" Denkaufgaben	-.46		-.46		.52
" Überschriften	-.43		-.47		.44
" Zahlenrätsel	-.37		-.32		.3o
" Wortverwendlung	-.31		-.38		.33
" Figurendrehung	-.54				.31
" Doppelverb			-.54		.35
" Figurenreihen	-.66				.46
" Buchstabenfolgen	-.55		-.31		.42
" Wortbedeutungen			-.61		.47
" Rechenarten	-.5o		-.34		.4o
" Satzteile	-.46		-.36		.36
" Würfelaufgaben	-.65				.47
" Wortanalogien	-.53		-.36		.42
" Bildreihen	-.39				.2o
Versuchsform IV, 1 Aufgabe 1	-.34	.3o	-.19	.12	.26
" " " 2A	-.13	.46	-.3o	-.13	.33
" " " 2Ø	-.16	.42	-.22	-.2o	.29
" " " 3A	-.12	.54	-.24	-.o6	.36
" " " 3Ø	-.11	.47	-.2o	-.o6	.28
" " " 4A	-.11	.51	-.28	-.o4	.35
" " " 4Ø	-.o4	.42	-.25	-.1o	.25
" " " 5A	-.o4	.62	-.24	-.o8	.46
" " " 5Ø	-.o2	.57	-.22	-.o9	.38
" " " 6A	-.27	.61	-.22	-.1o	.5o
" " " 6Ø	-.14	.46	-.14	-.11	.27
" " " 7A	.o1	.65	-.o4	.32	.53
" " " 7Ø	.o4	.62	-.o2	.36	.52
" " " 8A	.o2	.71	-.o1	.3o	.59
" " " 8Ø	.o9	.63	.o9	.19	.45
" " " 9A	-.o3	.71	.o5	.2o	.54
" " " 9Ø	-.o2	.57	.o9	.15	.35
" " " 1oA	-.15	.71	.o2	.o9	.54
" " " 1oØ	-.12	.67	-.o2	.12	.48
" " " 11A	-.12	.7o	.o9	.o3	.51
" " " 11Ø	-.o6	.68	.o7	-.o1	.47
% der aufgeklärten Varianz	26.23	3o.83	29.79	13.15	

*) Diese Faktorenlösung erfüllt das "Fürntratt-Kriterium" (s. Fürntratt 1969) nicht; der 4. Faktor ist nur durch 2 statt 3 Variablen mit $\frac{a^2}{h^2} > .5o$ gekennzeichnet. Dagegen wird bei der Faktoren-Lösung ohne die 1o Originalitätswerte auch der 4. Faktor durch 3 Variablen mit $\frac{a^2}{h^2} > .5o$ repräsentiert.

Die erste Varimax-Rotation mit 4 Faktoren erbrachte im Hinblick auf die Trennung des Faktors "divergentes Denken" von den übrigen Faktoren ein klares Bild.

Die Aufgaben der Versuchsform IV, 1 haben Faktorenladungen zwischen +.42 und +.71 auf dem 2. Faktor, Null- oder niedrige negative Ladungen auf dem ersten und Null- oder negative Ladungen bis .3o auf dem 3. Faktor.

Von diesem Ladungsmuster weicht nur die erste Aufgabe der Versuchsform IV, 1 insofern ab, als sie neben einer positiven Ladung von +.3o auf dem 2. eine ebenso hohe negative Ladung auf dem ersten Faktor aufweist.
Die Berechtigung, den 2. Faktor als Faktor des divergenten Denkens zu betrachten, ergibt sich nicht nur aus der Tatsache, daß sämtliche Aufgaben der Versuchsform IV, 1 mit Ausnahme der ersten ihre höchsten Ladungen auf diesem Faktor besitzen, sondern auch daraus, daß keine der übrigen Variablen ihre höchste Ladung auf diesem Faktor hat.

Im Hinblick auf die Ladungen auf dem 4. Faktor bieten die Aufgaben der Versuchsform IV, 1 ein weniger einheitliches Bild: Zeichnerisch zu lösende Aufgaben laden niedrig bis mäßig (.12 bis .36) auf dem 4. Faktor; auszunehmen ist hierbei die Aufgabe 4 ("Behausungen") mit Ladungszahlen von -.o4 für Produktivität und -.1o für Originalität. Die verbal zu beantwortenden Aufgaben laden um Null oder negativ (-.o6 bis -.2o) auf dem 4. Faktor; von diesem Ladungsmuster weichen 2 Aufgaben ab, die analog zu den Untertests von Wallach und Kogan "pattern meanings" und "line meanings" konstruiert worden sind und bei denen die kognitive Verarbeitung visueller Stimuli verlangt wird. Die Produktivitäts- und Originalitätsmaße dieser beiden Items laden zwischen .o1 und .12 auf dem 4. Faktor.

Wie bereits erwähnt, hat im Gegensatz zu den Aufgaben der Versuchsform IV, 1 keine der übrigen Variablen ihre höchste Ladung auf dem 2. Faktor. Die höchsten Ladungen der Zensuren entfallen auf den 3., die höchsten Ladungen des AST 4 auf den 3. und 4. Faktor. Die Untertests der beiden Intelligenztests haben ihre höchsten Ladungen auf dem 1., 3. und 4. Faktor.
Abgesehen von der Aufgabe 1 der Versuchsform IV, 1 ergibt sich die gleiche klare Trennung zwischen dem Faktor "divergentes Denken"

und den Faktoren, die für Noten, Schulleistungs- und Intelligenztestergebnisse bedeutsam sind, auch für Faktorenlösungen, bei denen 5, 6, 7 oder 8 Faktoren extrahiert wurden. Bei der 5-Faktoren-Lösung spaltet sich der Faktor "divergentes Denken" auf: ein Faktor zeigt die höchsten Ladungen bei den 4 rein verbalen Aufgaben ("verrückte Geschichten" und "lustige Vergleiche" mit je 2 Aufgaben), der andere Faktor hat die höchsten Ladungszahlen bei den zeichnerisch zu lösenden Aufgaben (außer Aufgabe 1) und bei den beiden Aufgaben "Bilderraten", die verbale Deutungen mehrdeutiger visueller Stimuli verlangen. Auch wenn 2 Faktoren des divergenten Denkens extrahiert werden, lädt keine der Schulleistungs- und Intelligenzvariablen auf einem dieser Faktoren bedeutsam; die Ladungen schwanken überwiegend um Null, die höchste Ladung von o.19 besitzt der Untertest "Wortschatz" des [illegible] auf dem Faktor "verbales divergentes Denken".

Die Lösungen mit 6, 7 und 8 Faktoren erbrachten eine weitere Aufspaltung des bzw. der Faktoren des divergenten Denkens; die klare Trennung zwischen den Faktorenladungsmustern der Aufgaben der Versuchsform IV, 1 und den Ladungsmustern aller übrigen Variablen zeigte sich auch hier.

Der Anteil der durch die jeweilige Faktorenlösung aufgeklärten Varianz an der Gesamtvarianz liegt zwischen 43 Prozent (bei 4 Faktoren) und 55 Prozent (bei 8 Faktoren).

Ein möglicher Einwand gegen die durchgeführten Faktorenanalysen könnte aus der relativ hohen Interkorrelation der Produktivitätsmit den Originalitätswerten der Aufgaben 2 bis 11 (die Koeffizienten liegen zwischen .61 und .83, Md =.76) abgeleitet werden; die beiden Maße sind insofern nicht voneinander unabhängig, als mit zunehmender Zahl der Antworten auf eine Frage die Wahrscheinlichkeit für eine statistisch seltene und somit originelle Antwort zunimmt. Die Eindeutigkeit des Faktors "divergentes Denken" könnte daher zumindest z.T. aus dieser Abhängigkeit der Originalitäts- von den Produktivitätswerten resultieren. Aus diesem Grund wurden sämtliche Daten mit Ausnahme der Originalitätspunkte ein zweites Mal faktoranalysiert.

Die Faktorenlösungen mit 3, 4 und 5 Faktoren erbrachten wieder einen Faktor "divergentes Denken", auf dem nur die Aufgaben der Versuchsform IV, 1 hohe Ladungen zeigten. Aufgabe 1 lud außerdem auf einem weiteren Faktor signifikant.
Die 6-Faktoren-Lösung erbrachte wieder die Aufspaltung in einen verbalen und einen figuralen Faktor des divergenten Denkens. Auf beiden Faktoren zeigten die Schulleistungs- und Intelligenztestvariablen mit einer einzigen Ausnahme Null- oder sehr niedrige Ladungen.

Der Prozentsatz der aufgeklärten Varianz an der Gesamtvarianz liegt bei diesen Faktorenanalysen zwischen 39 Prozent (3-Faktoren-Lösung) und 48 Prozent (6-Faktoren-Lösung).

Die Resultate der Faktorenanalysen können als Beleg für die faktorielle Validität der Aufgaben 2 - 11 der Versuchsform IV, 1 gewertet werden.

Insgesamt betrachtet sprechen die aufgrund verschiedener Methoden ermittelten Ergebnisse für eine positive Beantwortung der Frage, ob die Aufgaben der Versuchsform IV, 1 eine einheitliche Dimension intelligenten Verhaltens messen, die von gebräuchlichen Intelligenz- und Schulleistungstests nicht erfaßt wird.

Von dieser Feststellung ist lediglich die erste Aufgabe auszunehmen: sie korreliert mit allen Variablen einschließlich der Aufgaben der Versuchsform IV, 1 hochsignifikant (abgesehen von der Korrelation mit der Deutschnote, die nur auf dem 5 %-Niveau signifikant ist) und hat als einzige ein Ladungsmuster, das keine eindeutige Zuordnung ermöglicht.

Zu 2:

Zur Untersuchung der internen Konsistenz und Trennschärfe der Aufgaben der Versuchsform IV, 1 wurde das Programm DATANA[1)], das vor allem bei der Reliabilitäts- und Trennschärfenbestimmung von Fragebogenitems verwendet wird, eingesetzt,da es eine weitgehende Ausschöpfung der vorhandenen Informationen ermöglicht. Die Verfahren, die für Tests mit Mehrfach-Wahlaufgaben vorliegen, können dagegen nicht oder jedenfalls nur nach aufwendigen und mit Informationsverlust verbundenen Datentransformationen verwendet werden.

---

1) Programmautor: Dipl.Psych. K.U. Kohr,
Fachbereich o6 Psychologie der Universität Gießen

Das Programm erfordert eine Reduktion der Streubreite der Produktivitäts- und Originalitätspunkte auf 9 Kategorien, die durch Zusammenfassung von jeweils 2 oder 3 Rohpunkten zu einer Kategorie erreicht wurde.

Die nach DATANA berechneten Reliabilitätsmaße sind in Tabelle 9 wiedergegeben.

Tabelle 9: Reliabilitätskoeffizienten der 21 Produktivitäts- und Originalitätsmaße der Versuchsform IV, 1. (N = 526)

Schätzung der internen Konsistenz (Cronbach Alpha):	o.9o2
Odd-even split-half-Reliabilität:	o.896
Nach Spearman-Brown:	o.945
Nach Flanagan:	o.945
Nach Kristof:	o.946

Die Reliabilitätskoeffizienten erscheinen für ein aus 11 Items bestehendes Verfahren auch dann sehr hoch, wenn man berücksichtigt, daß diese Aufgaben ähnlich wie Untertests mit höheren Scores bewertet werden als Richtig-Falsch-Items.
Wenn man die Reliabilität des Verfahrens über die mittlere Aufgaben-Interkorrelation nach Formal 32 von Lienert (1967[2]) schätzt, erhält man einen Reliabilitätskoeffizienten von .85. (Da von 11 Aufgaben ausgegangen wird, basiert die Berechnung nur auf der gemittelten Korrelation der 11 Produktivitätswerte. Wenn man von allen 21 Testergebnissen und deren mittlerer Interkorrelation ausgeht, ergibt sich eine Reliabilitätsschätzung von .91.)

Tabelle 1o enthält die Trennschärfekoeffizienten der einzelnen Produktivitäts- und Originalitätswerte sowie die durchschnittliche Trennschärfe dieser Werte.

Tabelle 1o: Trennschärfekoeffizienten der 21 Items der Versuchsform IV, 1 (N = 526)

	Produktivität		Originalität	
Aufgabe 1	.433	.355 (korr.)	-	-
2	.53o	.469 "	.483	.421 (korr.)
3	.584	.526 "	.537	.473 "
4	.473	.422 "	.394	.335 "
5	.644	.598 "	.596	.545 "
6	.643	.596 "	.49o	.42o "
7	.663	.6o3 "	.624	.566 "
8	.7o3	.65o "	.591	.532 "
9	.696	.642 "	.615	.555 "
1o	.666	.625 "	.651	.592 "
11	.611	.569 "	.588	.527 "
Durchschnittliche Trennschärfe (unkorrigiert): o.588				
" " (korrigiert): o.53o				

Sowohl die durchschnittliche Trennschärfe als auch die Trennschärfekoeffizienten der einzelnen Produktivitäts- und Originalitätswerte liegen recht hoch; die niedrigsten Koeffizienten ergeben sich für Aufgabe 1 und Aufgabe 4.
Die Trennschärfe der Originalitätswerte ist deutlich niedriger als die der Produktivitätswerte.

Die Ergebnisse der Überprüfung der internen Konsistenz und der Trennschärfe der Aufgaben von IV, 1 können insgesamt als sehr gut bezeichnet werden.

Zu 3:

Die Hypothese, daß durch die zusätzliche Verwendung der Aufgaben der Versuchsform IV, 1 zu den Items des BBT die Vorhersage der Ergebnisse des Schulleistungstests AST 4 und der Zensuren nicht wesentlich verbessert werden kann, wurde bestätigt. Die durch eine schrittweise Regressionsanalyse[1] ausgewählten 4 besten Einzelprädiktoren waren immer Aufgaben des BBT; die durch sie erzielten multiplen

1) Programm REGT des Deutschen Rechenzentrums; Autor: Dr. F. Gebhardt, DRZ Darmstadt

Korrelationen konnten durch die Hinzunahme weiterer Prädiktoren nur unwesentlich erhöht werden. Tab. 11 zeigt die multiplen Korrelationen der ersten 6 Prädiktoren mit verschiedenen Schulleistungskriterien.

Tabelle 11: Multiple Korrelationen (R) der 6 besten Prädiktoren aus den Subtests des BBT und den Items der Versuchsform IV, 1 mit Schulleistungskriterien (N = 444)

Kriterium	Prädiktoren	R
AST 4 verbal	BBT-Untertest Satzteile	.54
"	BBT- " Denkaufgaben	.64
"	BBT- " Zahlenschrift	.68
"	BBT- " Doppelverb	.7o
"	Versuchsform IV, 1 Aufgabe 6 (Anzahl)	.71
"	" Aufgabe 8 (Originalität)	.71
AST 4 numerisch	BBT-Untertest Denkaufgaben	.58
"	BBT- " Zahlenreihen	.64
"	BBT- " Zahlenschrift	.67
"	BBT- " Figurenreihen	.68
"	BBT- " Doppelverb	.69
"	Versuchsform IV, 1 Aufgabe 1	.7o
"	" Aufgabe 1o (Originalität)	.7o
Deutschnote	BBT-Untertest Wortbedeutungen	.5o
"	BBT- " Zahlenschrift	.56
"	BBT- " Satzteile	.6o
"	BBT- " Denkaufgaben	.61
"	BBT- " Gegensätze	.62
"	Versuchsform IV, 1 Aufgabe 2 (Anzahl)	.62
Mathematiknote	BBT-Untertest Denkaufgaben	.52
"	BBT- " Zahlenreihen	.59
"	BBT- " Wortbedeutungen	.62
"	BBT- " Satzteile	.63
"	BBT- " Zahlenschrift	.64
"	Versuchsform IV, 1 Aufgabe 9 (Originalität)	.65
"	" Aufgabe 8 (Originalität)	.66
Fremdsprachennote	BBT-Untertest Wortbedeutungen	.43
"	BBT- " Satzteile	.52
"	BBT- " Denkaufgaben	.56
"	BBT- " Zahlenschrift	.58
"	BBT- " Doppelverb	.59
"	Versuchsform IV, 1 Aufgabe 4 (Anzahl)	.6o
"	" Aufgabe 3 (Anzahl)	.6o

Wenn statt der 11 Anzahl- und der 1o Originalitätswerte der Versuchsform IV, 1 die Gesamtwerte für Produktivität und Originalität zusammen mit den Untertests des BBT als Prädiktoren der 5 Schulleistungsmaße verwendet werden, befinden sich die beiden auf der Versuchsform IV, 1 basierenden Prädiktoren in keiner der 5 Regressionsanalysen unter den ersten 12 in die Regressionsgleichung eingehenden Variablen.

Aus den vorliegenden Ergebnissen kann also gefolgert werden, daß die durch die Versuchsform IV, 1 gemessene Fähigkeit für die Vorhersage der mit dem AST 4 und mit den Zensuren der 3 Hauptfächer erfaßten Schulleistungen ohne Bedeutung ist.

3.222 Untersuchungen an 4 Hauptschulklassen

Da abgesehen von den Ergebnissen zur faktoriellen Validität noch kaum Informationen über die Gültigkeit der Versuchsform IV, 1 als Test zur Erfassung des divergenten Denkens vorlagen, wurde in einer weiteren Voruntersuchung an 4 Klassen einer Hauptschule in Bad Dürkheim eine erste Untersuchung der Übereinstimmungsvalidität dieses Instruments durchgeführt.
Zu diesem Zweck wurden außer der Versuchsform IV, 1 folgende Verfahren eingesetzt:

a) Ein Lehrerfragebogen zum Unterrichtsverhalten der Schüler[1],
b) eine Schülerbefragung über das Verhalten aller Schüler in 3 kreativitätsfordernden Situationen,[2]
c) der Kreativitätstest von Wallach und Kogan.[3]

In den 4 fünften Klassen dieser Stichprobe war im Rahmen einer anderen Untersuchung eine Reihe zusätzlicher Daten erhoben worden, von denen folgende für die Beurteilung der Ergebnisse der Validitätsüberprüfung herangezogen wurden:
Beruf des Vaters (6 Stufen, in Anlehnung an Kleining und Moore 1968),
Deutschnote,
der Test "Zahlrechnen ZR 4+", ein diagnostischer Rechentest für 4. und höhere Klassen, der nach Fehlerzahl und -art ausgewertet wird,[4] und der "Kombinierte Lern- und Intelligenztest KLI 4+", dessen Teil L nach Aussage des Autors H. Schröder die Lernfähigkeit und dessen

---

1) Vgl. Anhang A 3

2) S. Anhang A 4

3) S. Anhang A 5. Die Aufgabeninhalte wurden unverändert von Wallach und Kogan übernommen, die Instruktionen wurden relativ frei übersetzt.

4) Zahlenrechnen ZR 4+ v.K. Krüger, E. Hylla und R. Bargemann aus der Reihe: Deutsche Schultests. Weinheim: Beltz 197o

Teil I die Intelligenz von Schülern 4. und 5. Klassen untersucht.

3.2221 Ergebnisse der Untersuchung mit dem Lehrerfragebogen

Der Lehrerfragebogen wurde vom Klassenlehrer ausgefüllt. Er enthielt in seiner ursprünglichen Fassung 13 fünfstufige Schätzskalen zum Unterrichtsverhalten der Schüler in alltäglichen schulischen Situationen. Die Auswahl dieser Situationen entsprach theoretischen Vorüberlegungen und/oder bereits vorliegenden Untersuchungsergebnissen zum Verhalten kreativer Schüler. Die Einschätzungen der Lehrer sollten Rückschlüsse ermöglichen über das Ausmaß, in dem die Schüler im Unterricht divergentes Denken und andere kreativitätsrelevante Verhaltensweisen an den Tag legen.

Gegen die Verwendung von Schätzskalen für Lehrer als Validitätskriterium eines Kreativitätstests kann eine ganze Reihe von Einwänden erhoben werden. Sie sollen im folgenden zusammen mit den Maßnahmen, die die Nachteile dieses Vorgehens beheben oder zumindest verringern sollten, kurz erörtert werden.

Aus den Ergebnissen der Untersuchung von Grote, Hajek u.a. (1969) wird deutlich, daß bei Lehrern nicht mit einer einheitlichen Auffassung des Begriffs "Kreativität" gerechnet werden kann.

Krause (1972), der die Lehrer in seiner Untersuchung den "Ideenreichtum" ihrer Schüler beurteilen ließ, fand eine Korrelation von .79 mit der von den Lehrern geschätzten Intelligenz; der Zusammenhang zwischen dem Ideenreichtum und den Ergebnissen des Kreativitätstests von Wallach und Kogan war dagegen wesentlich weniger eng (= .41).

Aufgrund dieser beiden Untersuchungsergebnisse erschien es nicht angebracht, das Lehrerurteil über die Kreativität der Schüler als Validitätskriterium zu verwenden; auch eine Erklärung oder Umschreibung des Begriffs Kreativität durch geläufigere Vokabeln wie Einfallsreichtum, Originalität u.ä. erschien angesichts des Ergebnisses von Krause nicht sinnvoll.

Eine bessere Validierungsmöglichkeit könnte darin gesehen werden, daß von den Lehrern Schätzurteile über Persönlichkeitsmerkmale der Schüler abgegeben werden, deren besonders starke oder besonders

schwache Ausprägung nach vorliegenden Untersuchungsergebnissen charakteristisch für kreative Schüler ist.[1] Die Gefahr dieses Verfahrens besteht darin, daß die Schätzungen häufig stärker von den Ansichten und Erwartungen der Beurteiler abhängen als von den zu beurteilenden Personen.

Überzeugend nachgewiesen wurde diese Urteilsverfälschung durch die "implizite Persönlichkeitstheorie" der Beurteiler vor allem von Hofer (1969), der in seiner Untersuchung u.a. Lehrer ihre Schüler auf einer Reihe von Eigenschaftsskalen einstufen ließ.

Um der Gefahr zu begegnen, daß das Validitätskriterium "Lehrerurteil" in der vorliegenden Untersuchung nur einen Teil der impliziten Persönlichkeitstheorie der Lehrer wiedergibt, wurden keine Eigenschaftsskalen verwendet, sondern es wurden konkrete Unterrichtssituationen vorgegeben, in denen der Lehrer das beobachtbare Verhalten seiner Schüler einschätzen sollte. Die Items fragten nicht danach, wie gut die Schüler irgend etwas machen, sondern wie gerne oder wie häufig sie sich in bestimmten vorgegebenen Situationen am Unterricht beteiligen oder wie oft sie ein bestimmtes Verhalten zeigen. Es wurde also nicht nach Fähigkeiten oder Leistungsniveau gefragt, sondern nach der Motivation für bestimmte Aktivitäten, weil es bei dieser Fragestellung leichter erschien, das beobachtbare Verhalten zur Grundlage der Beurteilungen zu machen, und weil angenommen wurde, daß die hierfür verwendeten Verhaltenskategorien weniger stark mit den Schulleistungen verknüpft sind als die Beurteilung bestimmter Unterrichtsaktivitäten nach irgendwelchen Gütekriterien. Die Möglichkeit, daß sich auch bei diesem Vorgehen Erwartungen der Lehrer über korrelierende Verhaltensweisen verfälschend auswirken, kann dadurch allerdings nicht ausgeschlossen werden; es wurde versucht, sie dadurch zu verringern, daß die Lehrer die in einer Zufalls-Reihenfolge aufgeführten Schüler itemweise beurteilten. Dieses Verfahren wurde der zweiten Möglichkeit, jeden Schüler nacheinander auf allen 13 Skalen einstufen zu lassen, vorgezogen, weil es weniger anfällig gegen solche Vermutungen über häufig zusammen auftretende Verhaltensweisen erschien.
Die Lehrer wurden nicht darüber informiert, daß die Ergebnisse des Fragebogens als Validitätskriterium für einen Kreativitäts-

---

1) S. dazu z.B. Cattel und Butcher 1968, Werner 1966, Werner und Bachtold 1969

test dienen sollten, da ihre Beurteilungen des Schülerverhaltens möglichst unbeeinflußt von ihren Ansichten über die intellektuellen Fähigkeiten der Schüler bleiben sollten.

Um die Bildung von Hypothesen über das eigentliche Ziel des Fragebogens zu erschweren und um Einflüsse durch Halo-Effekte möglichst auszuschalten, lag der sozial erwünschte Endpunkt nicht immer auf derselben Seite der Skalen. Außerdem wurden mehrere Situationen in den Fragebogen aufgenommen, bei denen eine aus der Sicht der Lehrer ungünstige Beurteilung als Indikator einer für kreative Schüler charakteristischen Motivation betrachtet wurde (Frage 1, 3, 8 und 11).

Die Fragen sollten folgende Verhaltensmerkmale kreativer Schüler ansprechen:

1. Mitarbeit und Engagement bei kreativitätsfordernder Unterrichtsgestaltung (Frage 2, 4, 6, 1o und 12),
2. geringes Interesse und wenig Einsatz bei Übungs- und Routinetätigkeiten (Frage 1 und 8),
3. diszipliniertes und an die Erfordernisse der Schulsituation angepaßtes Verhalten (Frage 3 und 11, eventuell Frage 9),
4. vom Lehrer und von den Klassenkameraden unabhängiges Verhalten (Frage 5,7, 9 und 13, eventuell auch Frage 11).

Aus den Untersuchungsergebnissen zur Persönlichkeitsstruktur kreativer Individuen[1)] wurden Hypothesen über die Faktorenstruktur des Fragebogens abgeleitet. Es wurde eine dreidimensionale Struktur mit folgenden Faktoren erwartet: Ein Faktor "Motivation für kreatives Verhalten" mit positiven Ladungen für die Fragen 2, 4, 6, 1o und 12 und eventuell negativen Ladungen für die Fragen 1 und 8, ein "Anpassungsfaktor", auf dem die Fragen 3, 11 und eventuell 9 laden (wobei theoretische Vorüberlegungen dafür sprachen, daß kreative Schüler sich nicht durch eine gute schulische Anpassung auszeichnen, während die Untersuchungsergebnisse von Grote, Hajek u.a. (1969) und von Krause (1972) die entsprechende Erwartung fraglich erscheinen ließen), und ein "Unabhängigkeitsfaktor", für den die Fragen

1) Einen Überblick über die in diesem Zusammenhang relevanten Ergebnisse gibt Mainberger 1975.

5, 7, 9 und 13 und eventuell auch die Frage 11 bedeutsam sein könnten.

Die Überprüfung dieser Hypothesen erfolgte mit einer Faktorenanalyse nach der Hauptachsen-Methode mit anschließender Varimax-Rotation.[1] Die Anzahl der beurteilten Schüler war 1o7, die Einschätzungen wurden von 4 Lehrern vorgenommen.

Ein Teil der Fragen wurde vor der Auswertung umgepolt, so daß für sämtliche Fragen ein hoher Punktwert als Indikator für kreatives Verhalten im Unterricht betrachtet werden konnte. Wegen der relativ geringen Variablenzahl wurden nur Lösungen mit 2,3 und 4 Faktoren gerechnet.
Die Eigenwertgrenze wurde auf o.5o festgesetzt. Keine der Faktorenlösungen erbrachte mehr als 2 durch mehrere Fragen repräsentierte Faktoren. Tabelle 12 zeigt die 3-Faktoren-Lösung, durch die 65 Prozent der Gesamtvarianz aufgeklärt werden.

Tabelle 12: Rotierte Faktorenmatrix der 13 Fragen des Lehrerfragebogens (N=1o7)

Frage	1. Faktor	2. Faktor	3. Faktor
1	-o.o2	-o.76	-o.o3
2	o.89	-o.12	-o.o1
3	o.15	-o.81	-o.2o
4	o.78	o.23	o.25
5	o.1o	o.55	o.1o
6	o.92	-o.o5	-o.1o
7	o.o5	o.22	o.o1
8	-o.o7	-o.77	o.1o
9	o.79	o.24	o.2o
1o	o.93	o.o3	-o.15
11	o.o4	-o.o6	-o.79
12	o.84	o.o4	-o.12
13	o.66	o.42	o.47
% der durch die 3 Faktoren aufgeklärten Varianz	58.o2	29.6o	12.38

1) Programm PAFA, s. dazu Anmerkung 1, S. 156

Der erste Faktor kann als die postulierte Dimension "Motivation für kreatives Verhalten" interpretiert werden; auf ihm laden die dieser Dimension hypothetisch zugeordneten Fragen sowie die Fragen 9 und 13, wobei Frage 13 insofern eine Sonderstellung hat, als sie als einzige auf mehr als einem Faktor substantielle Ladungen besitzt. Die erwarteten negativen Ladungen der Fragen 1 und 8 auf dem Faktor "Motivation für kreatives Verhalten" treten nicht auf.

Der zweite Faktor läßt sich in etwa als "Anpassung an die Schulsituation" interpretieren, er wird aber nicht wie erwartet aus den Fragen 3, 9 und 11 gebildet, sondern vor allem aus den Fragen 1, 3 und 8. Brave Schüler sind in den Augen der an der Untersuchung beteiligten Lehrer offenbar auch für Übungs- und Routine-Aufgaben und für Tätigkeiten, bei denen die saubere und exakte Ausführung im Vordergrund steht, gut motiviert (oder gilt gar die Umkehrung, daß Schüler, die nicht für alle schulischen Aufgaben motiviert sind, nicht "brav" sind?).

Der erwartete dritte Faktor erscheint in keiner der Faktorenlösungen mit der erforderlichen Deutlichkeit. Die Fragen nach unabhängigen Verhaltensweisen der Schüler waren die Items 5, 7, 9, 13 und - unter Vorbehalt - 11. Frage 5 lud am höchsten, aber mit umgekehrtem Vorzeichen auf dem Faktor "schulische Anpassung", was in etwa so interpretiert werden muß, daß Schüler, die die Ansichten ihrer Klassenkameraden übernehmen, in den Augen der 4 Lehrer weniger brav sind. Item 7, in dem nach der Unabhängigkeit vom Lehrer gefragt wird, lädt auf keinem der Faktoren substantiell.

Daß Frage 9 eine hohe Ladung auf dem ersten Faktor hat, erscheint bei Wegfall eines eindeutigen Unabhängigkeitsfaktors zwingend und bestätigt die Interpretation des ersten Faktors als "Motivation für kreatives Verhalten".

Frage 13 lädt am höchsten auf dem ersten und signifikant auf dem 2. und 3. Faktor. Aufgrund ihres inhaltlichen Zusammenhangs mit Frage 9 erscheint ihre Zugehörigkeit zum ersten Faktor plausibel.

Frage 11 schließlich hat eine hohe Ladung auf dem 3. Faktor, auf dem im übrigen nur noch die Frage 13 signifikant lädt, allerdings mit umgekehrtem Vorzeichen. Da somit der 3. Faktor kaum zu interpretieren ist, kann auch nichts darüber ausgesagt werden, welchen Verhaltensbereich der Schüler diese Frage in den Augen der Lehrer anspricht.

Aufgrund der dargestellten Ergebnisse und Überlegungen wurden für die Validitätsprüfung nur die Fragen 1, 2, 3, 4, 6, 8, 9, 1o, 12 und 13 des Lehrerfragebogens berücksichtigt. Sie wurden folgendermaßen zu 2 Verhaltenskategorien zusammengefaßt:
"Motivation für kreatives Verhalten", die durch die Fragen 2, 4, 6, 9, 1o, 12 und 13 untersucht wird, und "Anpassung an die Erfordernisse der derzeitigen Schulsituation", die durch die Fragen 1, 3 und 8 erfaßt wird.
Die für kreative Schüler wichtige Kategorie "unabhängiges Schülerverhalten" muß entfallen, wobei unklar bleibt, ob die verwendeten Fragen nicht geeignet waren, diesen Bereich zu erfassen, ob die Lehrer infolge einer eingeschränkten Wahrnehmung des Schülerverhaltens keine Aussagen über die Unabhängigkeit der Kinder machen können oder ob die derzeitige Schulsituation keine Manifestation unabhängigen Schülerverhaltens, das nicht zugleich mangelnde Anpassung an die Schulsituation bedeutet, zuläßt.

Da sich kreative Schüler durch ein hohes Maß an unabhängigem Verhalten auszeichnen, ist der um diesen Aspekt reduzierte Fragebogen weniger aussagekräftig. Das Validitätskriterium besteht nur noch aus dem Fragenkomplex "Motivation für kreatives Verhalten", da die Dimension "Anpassung an die Erfordernisse der derzeitigen Schulsituation" nicht als Außenkriterium für einen Kreativitätstest verwendet werden kann (s. S. 171).
Daß Hinweise auf eine positive Motivation für kreatives Verhalten als Indikator dafür betrachtet werden, daß dieses Verhalten auch tatsächlich auftritt - mit andern Worten: daß nicht das Verhalten selbst, sondern Anzeichen einer günstigen Motivation für dieses Verhalten das Validitätskriterium darstellen - bedarf einer kurzen Begründung, da die Fähigkeitskomponente hierbei vernachlässigt zu sein scheint.
Es wird davon ausgegangen, daß die Motivation für kreatives Verhalten als Folge positiver Erfahrungen bei Versuchen kreativer Problemlösungen entsteht. Positive Erfahrungen treten meist dann auf, wenn solche Versuche erfolgreich waren, wenn also Probleme kreativ bewältigt wurden, oder wenn die soziale Umwelt auf Bemühungen, Probleme kreativ zu lösen, unabhängig von Erfolg oder Mißerfolg positiv reagiert. Im ersten Fall ist ein direkter Schluß auf die Kreativität

des Problemlösers möglich, im zweiten kann mit den günstigen Auswirkungen einer kreativitätsfreundlichen Umwelt, die das vorhandene kreative Potential optimal fördert, gerechnet werden.

In einer Studie von ROOKEY (1973) wird eine Reihe von Untersuchungen referiert, die die hier dargestellte Annahme über die Beziehung zwischen Motivation und Fähigkeit zu kreativem Verhalten stützen. ROOKEY selbst fand in einer an 945 Schülern fünfter Klassen durchgeführten Untersuchung, daß eine positive Einstellung gegenüber kreativem Verhalten sowohl die ein Jahr später erneut untersuchte Einstellung als auch die ebenfalls nach einem Jahr gemessene Fähigkeit zu kreativem Verhalten vorhersagen konnte, während das Fähigkeitsmaß nur als Prädiktor für spätere Fähigkeitsmessungen verwendet werden konnte.

Für die Ratings der Lehrer zu dem Fragenkomplex "Motivation für kreatives Verhalten" wird eine mittlere Korrelation mit den Gesamtwerten für Anzahl und Originalität der Versuchsform IV, 1 erwartet. Wegen der vermutlich nicht sehr hohen Reliabilität des Lehrerurteils werden keine Korrelationen über .6o erwartet; bei Korrelationen unter .4o müßte die Validität der Versuchsform IV, 1 und/oder des Kriteriums "Lehrerurteil" als gering betrachtet werden.

Für die Beziehung des zweiten Fragenkomplexes "Anpassung an die Schulsituation" zu den Ergebnissen der Versuchsform IV, 1 wird wegen der Widersprüche zwischen theoretischen Vorüberlegungen und amerikanischen und deutschen Untersuchungsergebnissen keine Hypothese aufgestellt. Dieser Komplex wird in die Untersuchung aufgenommen, um zur Klärung des Zusammenhangs zwischen Kreativität und schulischer Anpassung beizutragen.

Ergebnisse der Korrelationsanalyse

Die Korrelationsanalyse[1)] wurde mit 149 Variablen, überwiegend Intelligenz-, Kreativitäts- und Verhaltensbeurteilungsresultaten, gerechnet.

Die für den ersten Gesamtwert des Lehrerfragebogens ("Motivation für kreatives Verhalten") und seine Beziehung zu den Kreativitätstestergebnissen relevanten Korrelationen werden in Tabelle 13 dargestellt.

---

1) Programm KOR 3 des Deutschen Rechenzentrums; Autor: D. Lörincz, DRZ Darmstadt

Tabelle 13: Produkt-Moment-Korrelationen ausgewählter Schülervariablen mit dem ersten Gesamtwert des Lehrerfragebogens (N=1o7)

	1. Gesamtwert der Lehrerfragen r
Geschlecht (männlich=1, weiblich= 2)[1]	- o.127
Vaterberuf (Stufen 1-6 nach Kleining und Moore; 1 = Oberschicht)	o.o85
Deutschnote	- o.148
ZR 4+ (Mathematik-Test, Auswertung nach Fehlerzahl)	- o.188
KLI-L (Kombinierter Lern- und Intelligenztest, Teil L)	o.o9o
KLI-I	o.192
Versuchsform IV, 1 Anzahl-Gesamtwert	o.18o
Versuchsform IV, 1 Originalitäts-Gesamtwert	o.151
Kreativitätstest von Wallach und Kogan, Anzahl-Gesamtwert	- o.o28
Kreativitätstest von Wallach und Kogan, Originalitäts-Gesamtwert	- o.o45
1. Schülerfrage	o.3o6 ss*
2. Schülerfrage	o.3o9 ss
3. Schülerfrage	o.282 ss
Gesamtwert der Schülerfragen	o.32o ss
2. Gesamtwert der Lehrerfragen (hohe Gesamtwerte bedeuten mangelnde Anpassung an die Schulsituation)	- o.136

*) Signifikante Korrelationskoeffizienten ($p \leq .o5$) sind durch s, sehr signifikante ($p \leq .o1$) durch ss gekennzeichnet.

1) Die an sich unzulässige Berechnung dieses Korrelationskoeffizienten soll nur ein Anhaltspunkt für geschlechtsspezifische Unterschiede in den Ergebnissen liefern; bei Berechnung der für diese Daten angemessenen punktbiserialen Korrelation würde sich allerdings der numerische Wert von r nicht ändern (s. dazu Lienert 1967[c], S. 98).

Der erste Gesamtwert der Lehrerfragen korreliert mit allen 148 Variablen dieser Untersuchung nur schwach, die meisten Korrelationen sind nicht signifikant. Die engsten Beziehungen bestehen zu den Ergebnissen der Schülerbefragung, was nicht verwunderlich ist, da die Beurteilungen der Lehrer und Schüler weitgehend von dem gleichen beobachtbaren Schülerverhalten ausgehen. Die Korrelationen sind allerdings nicht so hoch, daß eine Übernahme des Lehrerurteils durch die Schüler befürchtet werden müßte; diese Annahme kann bei Koeffizienten um o.3o als widerlegt betrachtet werden.

Die Beurteilung der "Motivation für kreatives Verhalten" korreliert mit den 4 Kreativitätsmaßen um Null bzw. nicht signifikant; für den Anzahl-Gesamtwert der Versuchsform IV, 1 ergibt sich der höchste Koeffizient, der aber ebenfalls nicht signifikant ist. In gleicher Höhe korreliert der erste Gesamtwert der Lehrerfragen mit dem Teil I des KLI und mit dem ZR 4+. Die Korrelation von -o.136 mit dem zweiten Gesamtwert des Lehrerfragebogens ist nicht signifikant; bei den an der Untersuchung beteiligten Lehrern besteht also kein gesicherter Zusammenhang zwischen ihrer Beurteilung der "Motivation für kreatives Verhalten" und ihrer Einschätzung der schulischen Anpassung der Kinder.

Die mit dem ersten Gesamtwert der Lehrerfragen berechneten Korrelationen lassen lediglich die Feststellung zu, daß die Validität der Versuchsform IV, 1 und/oder des Kriteriums Lehrerurteil gering ist.

Da die Korrelationen dieses Kriteriums mit dem Kreativitätstest von Wallach und Kogan, dessen Gültigkeit auch für eine Stichprobe bundesdeutscher Schüler durch eine von Krause (1972) durchgeführte Konstruktvalidierung bestätigt wurde, nicht von Null verschieden sind, besteht die Möglichkeit, daß nur das verwendete Validitätskriterium invalide ist.

Die Zusammenhänge zwischen dem zweiten Gesamtwert des Lehrerfragebogens ("Anpassung an die Erfordernisse der derzeitigen Schulsituation") und den ausgewählten Variablen werden in Tabelle 14 dargestellt.

Tabelle 14: Korrelationen ausgewählter Schülervariablen mit dem zweiten Gesamtwert des Lehrerfragebogens. (N=107)

	2. Gesamtwert der Lehrerfragen r
Geschlecht (männlich= 1, weiblich = 2)*	o.39o ss**
Vaterberuf***	o.o37
Deutschnote	- o.454 ss
ZR 4+ (Mathematik-Test, Auswertung nach Fehlerzahl)	- o.291 ss
KL1-L	o.231 s
KL1-I	o.2o3 s
Versuchsform IV, 1, Anzahl-Gesamtwert	o.272 ss
Versuchsform IV, 1, Originalitäts-Gesamtwert	o.2oo s
Kreativitätstest von Wallach und Kogan, Anzahl-Gesamtwert	o.228 s
Kreativitätstest von Wallach und Kogan, Originalitäts-Gesamtwert	o.192
1. Schülerfrage	o.398 ss
2. Schülerfrage	o.431 ss
3. Schülerfrage	o.143
Gesamtwert der Schülerfragen	o.353 ss

*) S. dazu Anmerkung 1, S. 172

**) Signifikante Korrelationskoeffizienten ($p \leq .05$) sind durch s, sehr signifikante ($p \leq .01$) durch ss gekennzeichnet.

***) S. dazu Tabelle 13

Zunächst fällt die Höhe des Zusammenhangs mit der Deutschnote auf, die ganz erheblich über der Korrelation der ersten Lehrerfrage mit dieser Note liegt. Eine gute Anpassung an die Erfordernisse der Schulsituation hat demnach für die 4 Lehrer mehr mit den Leistungen im Fach Deutsch zu tun als die Verhaltensweisen,die dem ersten

Fragenkomplex zugrundeliegen[1] und für die engere Zusammenhänge gerade mit der Deutschnote erwartet werden sollten. Als Erklärung für dieses überraschende Ergebnis sei auf die früher untersuchten Zusammenhänge zwischen Deutschzensur und Rechtschreibleistung verwiesen;[2] offenbar hat die Orthographie, eine Tätigkeit, für die die in den Fragen 1 und 8 angesprochenen Verhaltensweisen wahrscheinlich von Bedeutung sind, immer noch ein großes Gewicht bei der Beurteilung der Deutsch-Leistungen. Im übrigen ist anzunehmen, daß sich die schulische Anpassung gerade auf die Deutschzensur besonders stark auswirkt, da eine objektive Bewertung der Deutschleistungen infolge fehlender allgemeingültiger Beurteilungskriterien nicht möglich ist.[3]

Als zweites überrascht die hohe Korrelation zwischen der durch die Lehrerratings erfaßten schulischen Anpassung und dem Geschlecht der Schüler. Sie unterstreicht die Bedeutung der geschlechtsrollenspezifischen Sozialisation und bestätigt Hypothesen über die Gründe für die bessere Benotung der Mädchen[4]; Mädchen passen sich den Erfordernissen der Schulsituation entweder tatsächlich besser an oder die Mehrzahl der Lehrer unterliegt einem vom Geschlecht des Schülers abhängigen Beurteilungsfehler. In diesem Punkt läßt sich also ein durch die Befragung von 4 Lehrern erhaltenes Resultat durch die Ergebnisse anderer Untersuchungen bestätigen.

Die z.T. recht engen Zusammenhänge des Lehrerurteils mit den Schülerfragen wurden bereits bei der Darstellung der Ergebnisse des ersten Fragenkomplexes angesprochen; daß sie für den 2. Gesamtwert des Lehrerfragebogens eher noch enger sind, deutet entweder darauf hin, daß auch die Urteile der Schüler durch die gute soziale Anpassung ihrer Mitschüler beeinflußt wurden oder daß die Zuverlässigkeit der Lehrerratings bei dem 2. Fragenkomplex höher war. Daß die Fragen 1,3 und 8 Verhaltensweisen ansprechen, die für jede Art von Unterricht von Bedeutung sind, während die Items des ersten Fragenkomplexes z.T. nach Schülerverhalten fragen, das nur in einem kreativitätsfreundlichen Unterricht häufig beobachtet werden kann (z.B. Frage 2, 4, 6, 9), spricht für die zweite Erklärungsmöglichkeit.

---

1) Vgl.Anhang A 3
2) Kemmler 1970 , Weiß in Ingenkamp (Hrsg) 1971
3) S. z.B. Beck 1974
4) Vgl. z.B. Knoche 1969 und Starck 1974. Nach Starck haben die Jungen in Grund- und Hauptschulen, Realschulen und Gymnasien in allen Ländern der BRD höhere Sitzenbleiberquoten als die Mädchen.

Für den von den Lehrern beurteilten Grad der schulischen Anpassung war eine weniger enge Beziehung zu den 4 Kreativitätstestergebnissen erwartet worden als für die mit Hilfe der Lehrerratings ermittelte Stärke der Motivation für kreatives Verhalten. Das Gegenteil war der Fall: drei der vier Korrelationen der Kreativitätsmaße mit dem 2. Fragenkomplex sind signifikant oder sehr signifikant, das Originalitätsmaß von Wallach und Kogan ist fast signifikant (ab r =.195). Daß demgegenüber alle 4 Kreativitätstestergebnisse mit den Lehrerratings der "Motivation zu kreativem Verhalten" nicht korrelieren, verstärkt die Zweifel an der Validität des Lehrerurteils als Außenkriterium.
Deshalb kann aufgrund der mit dem Validitätskriterium "Lehrerurteil" erzielten Ergebnisse noch nichts über die Validität der Versuchsform IV, 1 ausgesagt werden.

Ergebnisse der Regressionsanalysen

Mit den beiden Gesamtwerten des Lehrerurteils wurden je 2 schrittweise Regressionsanalysen[1] gerechnet; Prädiktoren waren einmal die 11 Produktivitäts- und 1o Originalitätsmaße der Versuchsform IV, 1 und zum anderen die Ergebnisse des Kreativitätstests von Wallach und Kogan (39 Produktivitäts- und 39 Originalitätswerte). Die Ergebnisse der 4 Analysen werden in Tabelle 15 für die Gesamtzahl der in den Analysen berücksichtigten Prädiktoren und für eine ausgewählte Prädiktorenzahl dargestellt. Den Ausschlag für die Wahl einer bestimmten Anzahl von Prädiktorvariablen gaben die Ergebnisse der Berechnung der Schrumpfungskorrektur; die an zweiter Stelle aufgeführten "Populationsschätzwerte der multiplen Validitätskoeffizienten" ($\bar{R}$)[2] sind die jeweils höchsten Koeffizienten für die multiple Korrelation nach der Schrumpfungskorrektur. (Bei einer weiteren Vermehrung der Prädiktoren übertrifft die in der Population zu erwartende Schrumpfung der multiplen Korrelation die durch den zusätzlichen Prädiktor bewirkte Erhöhung von R.)

---

1) Programm REGT des Deutschen Rechenzentrums; Autor: Dr. F. Gebhardt, DRZ Darmstadt, bei Versuchsform IV, 1. Programm BMD O2R der Health Sciences Computing Facility, UCLA, bei den Testaufgaben von Wallach und Kogan. Da die maximale Anzahl der Prädiktorvariablen bei REGT 29 ist, konnte dieses Programm für die Aufgaben von Wallach und Kogan nicht verwendet werden; eine der Analysen wurde mit beiden Programmen durchgeführt, wobei die Ergebnisse trotz der unterschiedlichen Methoden - bei BMD o2 R Beginn mit dem besten Prädiktor ("stepwise solution" nach Kerlinger und Pedhazur 1973), bei REGT Beginn mit allen Prädiktoren und Eliminierung des jeweils schlechtesten ("backward solution") - weitgehend übereinstimmen.

2) Nach Lienert 1967[2], S. 418- 419

Tabelle 15: Multiple Korrelationen (R) und Populationsschätzwerte der multiplen Korrelationen ($\overline{R}$) der Produktivitäts- und Originalitätswerte zweier Kreativitätstests mit dem 1. und 2. Gesamtwert des Lehrerfragebogens (1. Motivation für kreatives Verhalten, 2. Anpassung an die Schulsituation). (N=122 bei der Versuchsform IV, 1, N=13o beim Kreativitätstest von Wallach und Kogan)

Kreativitätstest	1. Gesamtwert des Lehrerfragebogens		
	R	$\overline{R}$	Prädiktoren
Versuchsform IV, 1	.423	.13o	11 Anzahl-, 1o Originalitätswerte
	.367	.324	3 Ø, 4 Ø, 5 A, 5 Ø, 9 Ø*
Kreativitätstest von Wallach und Kogan	.765	.247	37 Anzahl-, 36 Originalitätswerte
	.69o	.6o2	12 Anzahl-, 12 Originalitätswerte
	2. Gesamtwert des Lehrerfragebogens		
	R	$\overline{R}$	Prädiktoren
Versuchsform IV, 1	.442	.189	alle Prädiktoren
	.4o6	.348	2 Ø, 3 A, 4 Ø, 7 Ø, 8 A, 11 A, 11 Ø
Kreativitätstest von Wallach und Kogan	.752	.254	33 Anzahl-, 37 Originalitätswerte
	.662	.563	1o Anzahl-, 14 Originalitätswerte

*) 3 Ø heißt Originalitätswert der Aufgabe 3,
5 A heißt Anzahl der akzeptierten Antworten zu Aufgabe 5

Der Kreativitätstest von Wallach und Kogan mit seinen 39 Anzahl- und 39 Originalitätswerten sagt beide Kriterien erheblich besser vorher als die Versuchsform IV, 1 mit insgesamt 21 Prädiktoren. Zwar fallen bei Einbeziehung aller 73 bzw. 7o Prädiktoren die korrigierten multiplen Korrelationen sehr stark ab, aber auch die mit jeweils 24 Prädiktoren errechneten multiplen Korrelationskoeffizienten liegen höher

als diejenigen der 21 Anzahl- und Originalitätswerte der Versuchsform IV, 1, da sie auf einer optimalen Auswahl aus 78 Prädiktorvariablen basieren.

Beide Tests sagen die für kreatives Verhalten unterschiedlich bedeutsamen Kriterien etwa gleich gut voraus.

Wegen der zweifelhaften Validität des Außenkriteriums "Lehrerurteil" (s. dazu S. 171-176) können auch die Ergebnisse der dargestellten Regressionsanalysen keine Entscheidungshilfen für die weitere Arbeit mit der Versuchsform IV, 1 bieten.

3.2222 Ergebnisse der Schülerbefragung

Die zweite Methode der Validitätsüberprüfung bestand darin, die Kinder selbst das Verhalten aller Schüler ihrer Klasse in drei verschiedenen kreativitätsfordernden Situationen einstufen zu lassen. Es wurde erwartet, daß die drei Mittelwerte für jeden Schüler, die aus den Einstufungen aller Schüler für jede dieser Situationen errechnet wurden, Rückschlüsse auf das Ausmaß kreativen Verhaltens ermöglichen und somit als Validitätskriterium verwendet werden können.

Die größte Schwierigkeit bei der Konstruktion der Schätzskalen für die Schüler bestand darin, Fragen zu finden, für die aufgrund des täglichen Umgangs mit den Klassenkameraden und der Beobachtungen während des Unterrichts eine Beurteilungsgrundlage gegeben ist, ohne daß einfach nach bestimmten Verhaltensweisen im Unterricht gefragt wird. Letzteres erschien wegen der Gefahr, daß die Ratings der Schüler vom Lehrerurteil in Form von Zensuren, Lob und Tadel stark beeinflußt werden, nicht angemessen.

Da im Unterricht selten ausgesprochen kreative Aktivitäten verlangt werden, haben die Schüler kaum Gelegenheit, ihre Klassenkameraden in kreativitätsfordernden Situationen zu beobachten.[1)] Die Fragen mußten also von fiktiven Situationen ausgehen; wenn sie trotzdem für alle Schüler interessant sein sollten, mußten diese Situationen grundsätzlich im Bereich des Möglichen liegen und möglichst konkret geschildert werden. Diese Forderungen wurden erfüllt durch die Begründung der Fragen mit der Planung eines Landschulheim-Aufenthalts.

Um auch die 5 Einstufungskategorien möglichst konkret zu formulieren, wurden nicht wie bei dem Lehrerfragebogen Zahlen, sondern verbale

---

1) So wurde z.B. in einer kleinen Voruntersuchung die Frage, welche Mitschüler gut Spiele erfinden können, von einem Fünftkläßler mit der Gegenfrage beantwortet, woher er das wissen solle, in der Schule würden sie nie spielen.

Beschreibungen der Kategorien verwendet ("... Schüler,die das ganz besonders gut können", "... Schüler, die das auch noch recht gut können" etc.).

Da die Schülerzahl der an dieser Untersuchung beteiligten Klassen über 3o lag, wurden von jedem Schüler mehr als 3 x 3o Schätzurteile verlangt. Um Konzentration und Ausdauer der Kinder nicht unnötig zu strapazieren, wurden die drei Fragen an drei verschiedenen Tagen gestellt. Auf diese Weise wurde auch eine übereinstimmende Einstufung der Klassenkameraden auf allen 3 Skalen erschwert.

Schließlich sollte verhindert werden, daß die Kinder die Einstufungen zu schnell und unüberlegt vornehmen; diese Gefahr bestand vor allem für die zweite und dritte Frage als Folge von Langeweile und Überdruß. Deshalb wurde statt der üblichen Ankreuzungen der jeweils für zutreffend erachteten Kategorie ein Verfahren gewählt, das eine mehr spielerische Manipulation von schmalen Papierstreifen, auf denen die Namen aller Schüler der betreffenden Klasse standen, erforderte. Jeder Schüler erhielt an jedem der 3 Untersuchungstage eine kleine Pappschachtel, in der alle Namensstreifen lagen, und einen Bogen Papier, der so gefaltet und geheftet war, daß 5 Taschen entstanden. Auf diesen Taschen standen die bereits erwähnten 5 verbal charakterisierten Kategorien.

Die Schüler holten die Namensstreifen in beliebiger Reihenfolge aus den Pappschachteln und schoben sie in die nach ihrer Ansicht passende Tasche. Auf dem rechten Rand der Blätter stand in der Mitte die jeweilige Frage, darüber schrieben die Kinder ihren Namen.

Der freie Blattrand wurde nach Abschluß des Zuordnungsvorgangs über die Taschen geklappt, so daß die Namensstreifen nicht herausrutschen konnten.[1)]

Für die Auswertung wurden die verbal formulierten Kategorien durch die Zahlen 1 - 5 ersetzt und die Einstufungen für jeden Schüler aufsummiert. Da bei der Manipulation der Namensstreifen gelegentlich ein Schüler nicht zugeordnet wurde, so daß für eine kleinere Anzahl von Schülern eine Einstufung zu wenig vorlag, wurden die Summenwerte durch die Anzahl der Einstufungen dividiert; dadurch wurde auch erreicht, daß diese Mittelwerte über die 4 Klassen hinweg vergleichbar waren.

---

1) S. Anhang A 4

Wenn die Mittelwerte der Schätzurteile zu den drei inhaltlich ganz verschiedenen Fragen als Validitätskriterien für einen Kreativitätstest dienen sollten, mußte die Höhe ihrer Interkorrelationen überprüft werden. Die Korrelationskoeffizienten sind in Tabelle 16 wiedergegeben.

Tabelle 16: Produkt-Moment-Korrelationen der 3 Mittelwerte der Schülerbefragung (N=1o7)

	Frage 1 (Vorschläge für Bastelarbeit machen)	Frage 2 (Theaterstück ausdenken)
Frage 2	$r_{12} = o.937$	-
Frage 3 (lustige Streiche ausdenken)	$r_{13} = o.79o$	$r_{23} = o.792$

Obwohl die Schüler darauf hingewiesen worden waren, daß sie sich bei den Einstufungen nicht an den Schulleistungen orientieren sollen und daß infolge der Verschiedenheit der Programmpunkte wahrscheinlich verschiedene Schüler die Gestaltung der drei Abende übernehmen sollten, korrelierten die Ergebnisse der Fragen sehr hoch miteinander.

Die etwas niedrigeren Korrelationen mit der dritten Frage sind eventuell darauf zurückzuführen, daß die Fähigkeit der Mädchen, sich "lustige Streiche auszudenken", schlechter beurteilt wurde als diejenige der Jungen. Eine Überprüfung der Ergebnisse der Schülerbefragung mit dem t-Test für unabhängige Stichproben ergab nur für die dritte Frage einen hochsignifikanten t-Wert. ($p < .oo1$).[1)]

1) Ob die Voraussetzungen für die Anwendung eines parametrischen Prüfverfahrens gegeben sind, wurde nicht überprüft. Da der Unterschied zwischen den Beurteilungen der Jungen und Mädchen auch bei Anwendung eines parameterfreien Verfahrens (U-Test von Mann-Whitney) ein hochsignifikantes Ergebnis erbrachte ($Z > 4$), wurden die folgenden klasseninternen Überprüfungen, deren Ergebnisse nicht weiterverwendet wurden, mit dem schneller durchführbaren t-Test vorgenommen.

Trotz dieser geschlechtsspezifischen Beurteilungsunterschiede[1] erschien es aufgrund der hohen Interkorrelationen gerechtfertigt, die drei Ergebnisse der Schülerbefragung zu einem Gesamtwert zusammenzufassen. Eine Überprüfung der Gesamtwerte der Jungen und Mädchen aller 4 Klassen erbrachte keinen signifikanten Unterschied ($t = 1{,}63$, $df = 133$, $p > 0.5$ bei einseitigem Testen).

Ergebnisse der Korrelationsanalyse

Für die aus der Schülerbefragung gewonnenen Außenkriterien wurden engere Zusammenhänge mit den Kreativitätstestergebnissen erwartet als für die beiden Gesamtwerte der Lehrerurteile, da die Reliabilität der aus den Ratings aller Schüler berechneten Mittelwerte wesentlich höher sein dürfte als die der Urteile eines einzelnen Lehrers.

Korrelationen der Schülerratings mit den Testresultaten von .40 und darüber sollten als Hinweise auf die Validität der Versuchsform IV, 1 gewertet werden.

Die Korrelationen einiger Variablen mit den Ergebnissen der drei Schülerfragen und dem Gesamtwert aus diesen drei Fragen sind in Tabelle 17 dargestellt.

---

1) Über die Gründe für die unterschiedliche Beurteilung der Jungen und Mädchen können nur Vermutungen angestellt werden. Grundsätzlich sind zumindest drei Erklärungen in Betracht zu ziehen:

1. Jungen können sich tatsächlich mehr "lustige Streiche" ausdenken als Mädchen.
2. Die geschlechtsspezifische Erziehung in unserer Gesellschaft führt zu bestimmten Erwartungen, die eine Überschätzung der Jungen und/oder eine Unterschätzung der Mädchen auf diesem Gebiet zur Folge haben.
3. Die Beurteilungsunterschiede werden nicht durch generelle Erziehungseinflüsse verursacht, sondern gehen auf klassenspezifische Beeinflussungen durch bestimmte Lehrer zurück.

Eine Entscheidung über diese und eventuell weitere Erklärungsmöglichkeiten kann aufgrund der vorliegenden Daten nicht getroffen werden. Immerhin wurden für die 4 Klassen derselben Schule Beurteilungsunterschiede gefunden: Die Gesamtmittelwerte der 3 Fragen zeigten in einer Klasse eine signifikante geschlechtsspezifische Differenz zu Gunsten der Jungen ($p < .05$), in den anderen Klassen waren die Mittelwertunterschiede - einmal zu Gunsten der Mädchen, zweimal zu Gunsten der Jungen - nicht signifikant (Prüfung mit dem t-Test für unabhängige Stichproben; s. dazu Anmerkung 1, S. 180).
Bei der 3. Frage waren die Beurteilungen der Jungen aus den Klassen 5a und 5c hochsignifikant besser ($p < .01$ und $p < .001$) als die der Mädchen, die Jungen der Klasse 5b wurden signifikant besser eingestuft als die Mädchen ($p < .05$) und in der Klasse 5d wurden Jungen und Mädchen gleich gut beurteilt ($\bar{X}_J = 329$, $\bar{X}_M = 328$).

Tabelle 17: Korrelationen ausgewählter Schülervariablen mit den Ergebnissen dreier Schülerratings und dem daraus gebildeten Gesamtwert (N = 107)

	1. Schülerfrage (Vorschläge für Bastelarbeit)	2. Schülerfrage (Theaterstück ausdenken)	3. Schülerfrage (lustige Streiche ausdenken)	Gesamtwert der Schülerfragen
Geschlecht ** (männl. = 1) weibl. = 2)	.032	.033	- .312 ss*	- .074
Vaterberuf**	- .016	- .010	- .076	- .034
Deutschnote	- .584 ss	- .611 ss	- .443 ss	- .581 ss
ZR 4+	- .506 ss	- .506 ss	- .367 ss	- .489 ss
KLI-L	.256 ss	.283 ss	.231 s	.271 ss
KLI-I	.466 ss	.468 ss	.423 ss	.476 ss
IV, 1 Anzahl -Gesamtwert	.430 ss	.526 ss	.319 ss	.451 ss
IV, 1 Originalitäts-Gesamtwert	.377 ss	.430 ss	.248 s	.375 ss
Wallach und Kogan, Anzahl	.326 ss	.365 ss	.254 ss	.332 ss
Wallach und Kogan, Originalität	.333 ss	.355 ss	.244 s	.329 ss
1. Gesamtwert der Lehrerfragen	.306 ss	.309 ss	.282 ss	.320 ss
2. Gesamtwert der Lehrerfragen	- .398 ss	- .430 ss	- .143	- .353 ss

* Signifikante Korrelationskoeffizienten ($p \leq .05$) sind durch s, sehr signifikante ($p \leq .01$) durch ss gekennzeichnet.

** S.dazu Tabelle 13

Die Korrelationen der Ergebnisse der Versuchsform IV, 1 mit denjenigen der Schülerbefragung sind überwiegend hochsignifikant und durchweg höher als diejenigen der Testergebnisse mit den Lehrerurteilen, aber sie sind zum Teil weniger hoch als erwartet (vgl. S.181). Außerdem korreliert das Validitätskriterium höher mit der Deutschnote und dem Ergebnis eines Rechentests als mit den Ergebnissen der beiden Kreativitätstests. Dieses Resultat weist entweder darauf hin,

daß die Urteile der befragten Schüler über ihre Mitschüler zu einem nicht unwesentlichen Teil von den Schulleistungen bestimmt wurden oder daß sich die Fähigkeit der Schüler zu divergentem Denken sowohl auf die Schülerratings als auch auf die Schulleistungen positiv auswirkt. Die beiden Interpretationsmöglichkeiten schließen sich allerdings nicht aus; es spricht sogar einiges dafür, daß beide gemeinsam die Ergebnisse am besten erklären.(So besteht z.B. eine ebenfalls hochsignifikante Beziehung zwischen dem Anzahl-Gesamtwert der Versuchsform IV, 1 und der Deutschnote von -.403; nach Auspartialisierung des Einflusses der Deutschnote fällt die Korrelation zwischen dem Ergebnis des Kreativitätstests und demjenigen der 2. Schülerfrage von .526 auf .386 relativ stark ab, aber die verbleibende Partialkorrelation ist immer noch recht hoch.)

Ein weiteres Schülermerkmal, das mit den Resultaten der Schülerbefragung relativ eng zusammenhängt, ist die Intelligenz. Die Ergebnisse des KLI, Teil I, korrelieren mit den Schülerratings z.T. ebenso hoch oder höher als die Testwerte der Versuchsform IV, 1. Nach Auspartialisierung der mit dem KLI gemessenen Intelligenz sinkt die Korrelation des Anzahl-Gesamtwertes mit den Ratings zur 1. Schülerfrage von .430 auf .350, mit den Ratings zur 2. Schülerfrage von .526 auf .464 und mit denjenigen zur 3. Schülerfrage von .319 auf .228 ab, was bedeutet, daß die Schülerurteile nicht allzu stark von der Intelligenz beeinflußt wurden.

Sämtliche Resultate der Schülerbefragung korrelieren mit den Ergebnissen der Versuchsform IV, 1 höher als mit denjenigen des Kreativitätstests von Wallach und Kogan. Dies könnte allerdings nur bei nachgewiesener Validität des Außenkriteriums als Hinweis auf die höhere Validität der Versuchsform IV, 1 gewertet werden.
Im Gegensatz zu den Ergebnissen des Tests von Wallach und Kogan sind die Korrelationen der Schülerratings mit den Originalitäts-Gesamtwerten der Versuchsform IV, 1 durchweg niedriger als diejenigen mit den Anzahl-Gesamtwerten. Eine mögliche Erklärung hierfür ist, daß die gesamte Auswertung des Tests von Wallach und Kogan von einem Auswerter durchgeführt wurde, während bei der Versuchsform IV, 1 mehrere Personen an der Originalitätsauswertung beteiligt waren.

Trotz der durchweg signifikanten und im Vergleich mit den Ergebnissen des Lehrerfragebogens erheblich höheren Korrelationen der Kreativitäts-

testergebnisse mit den Schülerratings ist es außerordentlich schwierig, die Validität dieser zweiten Gruppe von Außenkriterien zu beurteilen. Abgesehen von der Zirkularität des Schlusses von einer hohen Korrelation zwischen den Ergebnissen eines noch nicht validierten Tests und einem für die Validierung dieses Tests entwickelten "subjektiven" Kriteriums[1] sind die Validitätskoeffizienten doch etwas niedriger und die Beziehungen zwischen den Schülerratings und den Schulleistungsvariablen etwas höher als erwartet. Mit Sicherheit kann nur gesagt werden, daß die Validität der Außenkriterien "Schülerratings" in dieser Untersuchung höher ist als diejenige des Außenkriteriums, das aus den Ergebnissen des Lehrerfragebogens entwickelt wurde.

Wegen der nicht genügend abgesicherten Validität der aus der Schülerbefragung gewonnenen Außenkriterien erscheint eine Beurteilung der Validität der Versuchsform IV, 1 aufgrund der Ergebnisse der Korrelationsanalyse nicht möglich.

Ergebnisse der Regressionsanalysen

Mit den Ergebnissen der 3 Schülerfragen und dem daraus gebildeten Gesamtwert wurden je 2 schrittweise Regressionsanalysen gerechnet; Prädiktoren waren einmal die Anzahl- und Originalitätswerte der Items der Versuchsform IV, 1, zum anderen die ebenfalls nach Produktivität und Originalität ausgewerteten Ergebnisse des Kreativitätstests von Wallach und Kogan.[2]

Tabelle 18 zeigt die multiplen Korrelationskoeffizienten für die Gesamtzahl der Items der Versuchsform IV, 1 und des Kreativitätstests von Wallach und Kogan[3] und für je eine optimale Auswahl von Testitems (vgl. dazu S.176) sowie die Populationsschätzwerte der multiplen Korrelationen.

---

1) S. dazu Lienert 1967², S. 17-18 und S. 274-279, der diese Problematik nicht anspricht. (S. 17: "Dabei wird das Kriterium als ausreichend valide angenommen, und zwar in dem Maße, in dem es reliabel, also reproduzierbar ist.")

2) Die verwendeten Programme sind in Anmerkung 1, S. 176, kurz beschrieben.

3) Die 78 Anzahl- und Originalitätswerte des Tests von Wallach und Kogan wurden in den schrittweisen Regressionsanalysen nicht alle berücksichtigt, da der für die Aufnahme der letzten Variablen erforderliche F-Wert ($p < .01$) nicht erreicht wurde; die Zahl der Prädiktorvariablen liegt zwischen 7o und 73.

Tabelle 18: Multiple Korrelationen (R) und Populationsschätzwerte der multiplen Korrelationen ($\bar{R}$) der Anzahl- und Originalitätswerte zweier Kreativitätstests mit den Ergebnissen der Schülerbefragung. (N=122 bei Versuchsform IV, 1; N=13o bei dem Test von Wallach und Kogan)

Kreativitätstest	1. Schülerfrage		
	R	$\bar{R}$	Prädiktoren
Versuchsform IV, 1	.586	.463	alle Testscores (11 Anzahl-, 1o Originalitätswerte)
	.572	.54o	1, 3 A, 3 Ø, 4 Ø, 6 A, 8 Ø, 9 A
Kreativitäts-test von Wallach und Kogan	.79o	.4o6	36 Anzahl-, 36 Originalitätswerte
	.721	.617	13 Anzahl-, 17 Originalitätswerte
	2. Schülerfrage		
	R	$\bar{R}$	Prädiktoren
Versuchsform IV, 1	.623	.516	alle 21 Testscores
	.596	.573	1, 3 A, 3 Ø, 4 A, 8 Ø, 1o Ø
Kreativitäts-test von Wallach und Kogan	.788	.376	37 Anzahl-, 36 Originalitätswerte
	.722	.6o2	17 Anzahl-, 16 Originalitätswerte
	3. Schülerfrage		
	R	$\bar{R}$	Prädiktoren
Versuchsform IV, 1	.573	.442	alle 21 Testscores
	.553	.5oo	1, 3 A, 3 Ø, 4 Ø, 5 A, 6 A, 8 Ø, 9 A, 11 A, 11 Ø
Kreativitäts-test von Wallach und Kogan	.765	.329	34 Anzahl-, 36 Originalitätswerte
	.715	.6o2	15 Anzahl-, 16 Originalitätswerte
	Gesamtwert der Schülerfragen		
	R	$\bar{R}$	Prädiktoren
Versuchsform IV, 1	.6o2	.485	alle 21 Testscores
	.58o	.55o	1, 3 A, 3 Ø, 4 Ø, 6 A, 8 Ø, 9 A
Kreativitäts-test von Wallach und Kogan	.78o	.397	35 Anzahl-, 35 Originalitätswerte
	.7o2	.618	13 Anzahl-, 11 Originalitätswerte

Die Ergebnisse der Schülerratings werden durch die Items des Kreativitätstests von Wallach und Kogan nur zum Teil besser vorhergesagt als durch die Aufgaben der Versuchsform IV, 1. Nach Durchführung der Schrumpfungskorrektur sagt die Versuchsform IV, 1 die 4 Kriterien besser vorher als der Test von Wallach und Kogan. Wenn nicht alle, sondern eine optimale Anzahl von Prädiktoren verwendet werden, ergeben sich für den Test von Wallach und Kogan die besseren Resultate.

Die Ergebnisse der Schülerbefragung werden durch beide Tests besser vorhergesagt als die beiden Gesamtwerte des Lehrerfragebogens; dies gilt für die Versuchsform IV, 1 in stärkerem Ausmaß als für den Kreativitätstest von Wallach und Kogan.

Die im Vergleich mit den Lehrerurteilen bessere Vorhersagbarkeit der Schülerratings durch die Kreativitätstestergebnisse kann natürlich nicht als Beleg für die höhere Validität dieser Außenkriterien betrachtet werden. Es kann lediglich gesagt werden, daß die Resultate der Regressionsanalysen der aus den Ergebnissen der Korrelationsanalyse abgeleiteten günstigeren Beurteilung der Außenkriterien "Schülerratings" nicht widersprechen.

Bei einer vergleichenden Wertung der Außenkriterien "Lehrerurteil" und "Schülerbefragung" im Hinblick auf ihre Verwendung in der Hauptuntersuchung muß allerdings berücksichtigt werden, daß ersteres auf den Einschätzungen je eines Lehrers basiert, während die aus den Schülerratings berechneten Außenkriterien Mittelwerte aus den Einstufungen der Schüler einer ganzen Klasse sind. Ein die Ergebnisse verfälschender Stichprobenfehler ist also bei den Beurteilungen der 4 beteiligten Lehrer wahrscheinlicher als bei den Schülerratings. Auch ein schulspezifischer Effekt ist bei dieser auf eine Schule beschränkten Voruntersuchung für die Resultate des Lehrerfragebogens eher zu befürchten als für diejenigen der Schülerbefragung, da die Schüler dieser 4 fünften Klassen diese Schule zum Zeitpunkt der Untersuchung erst seit etwa 1o Monaten besuchten. Diese beiden eventuellen Fehlerquellen könnten in einer eine größere Anzahl von Lehrern und Schulen umfassenden Untersuchung vermieden werden. Ein abschließendes Urteil über die Verwendbarkeit des Lehrerfragebogens und der Schülerratings als Außenkriterien für einen Kreativitätstest dürfte also erst nach einer solchen Untersuchung möglich sein.

3.2223 Ergebnisse der Untersuchung mit dem Kreativitätstest von Wallach und Kogan

Außer den beiden dargestellten Versuchen einer äußeren Validierung der Versuchsform IV, 1 wurde auch die Möglichkeit einer inneren Validierung mit Hilfe des Kreativitätstests von Wallach und Kogan überprüft. Die Einbeziehung eines zweiten Kreativitätstests in die Validierungsuntersuchungen bot außerdem eine zusätzliche Möglichkeit der Konstruktvalidierung durch einen Vergleich der konvergenten und der diskriminanten Validität der Verfahren.[1] Für diesen Test liegen eine Reihe amerikanischer und deutscher Untersuchungen an jüngeren Schülern vor,[2] deren Ergebnisse z.T. als Validitätsnachweise betrachtet werden können. Die Hauptgründe für seine Verwendung in dieser Untersuchung waren, daß die einzelnen Aufgaben untereinander einigermaßen hoch und mit Intelligenztestaufgaben erheblich niedriger korrelieren und daß bereits eine Reihe von Testergebnissen für bundesdeutsche Schüler dieser Altersstufe vorliegt.

Obwohl Wallach und Kogan großen Wert auf Individualtestung ohne Zeitbegrenzung und mit spielerischer Atmosphäre legen, mußte der Test klassenweise und mit einer - allerdings großzügig bemessenen - Zeitbegrenzung durchgeführt werden, da die beteiligte Schule für diese dritte Untersuchung nicht mehr allzu viel Zeit zur Verfügung stellen konnte. Die Testdurchführung dauerte insgesamt etwa 2 volle Zeitstunden und war auf 2 verschiedene Schultage verteilt.[3] Aufgrund einer Reihe von Untersuchungsergebnissen konnten keine sehr hohen Korrelationen zwischen den Ergebnissen der Versuchsform IV, 1 und denjenigen des Kreativitätstests von Wallach und Kogan erwartet werden.

Wallach und Kogan (1965) begründen ihre Kritik an den bis zu diesem Zeitpunkt vorliegenden Kreativitätstests mit den geringen Interkorrelationen der Aufgaben dieser Verfahren und der zumindest ebenso hohen Korrelation der Kreativitäts- und Intelligenztestitems; der Vorwurf der mangelnden inneren Konsistenz der zu einem "Kreativitätstest" zusammengefaßten Aufgaben trifft vor allem die Untersuchungen von Getzels und Jackson (1962), Cline, Richards und Abe (1962), Cline, Richards und Needham (1963), Flescher (1963), Guilford und Christensen (1956) und Wilson, Guilford, Christensen und Lewis (1954). Die beiden Autoren setzten sich das Ziel, einen Kreativitätstest zu konstruieren, dessen Aufgaben mit Intelligenztest-

---

1) S. dazu Campbell und Fiske 1959
2) Wallach and Kogan 1965, Kogan and Morgan 1969, Kogan and Pankove 1971, Ward 1968, Krause 1972, Grote, Hajek u.a. 1969. S. dazu S. 77ff
3) Anhang A enthält ein Testexemplar; die allgemeinen Durchführungshinweise waren die gleichen wie bei Versuchsform IV, 1.

items nicht, die jedoch untereinander hoch korrelieren. Die 5 Untertests ihres Verfahrens korrelierten mit den Ergebnissen zweier Intelligenz- und eines Schulleistungstests tatsächlich um Null bzw. insignifikant, aber die durchschnittliche Korrelation der 5 Produktivitäts- und der 5 Originalitätswerte dieser Untertests beträgt nur .41. Wallach und Kogan gingen zwar von 45 Korrelationskoeffizienten aus, sie bezogen also die Koeffizienten für den Zusammenhang zwischen "Produktivität" und "Originalität" bei derselben Aufgabe in die Durchschnittsberechnung ein, aber sie trafen Vorkehrungen, um eine unangemessene Erhöhung der Korrelationen zu verhindern.[1] Krause (1972) erhält bei etwas abweichendem Vorgehen[1] ebenfalls einen Koeffizienten von .41.[2]

Wenn trotz der Zielsetzung, einen Test mit homogenen Aufgaben zu entwickeln, und trotz der die innere Konsistenz der Ergebnisse begünstigenden Testsituation (Einzeltestung in spielerischer Atmosphäre ohne Zeitbegrenzung bei Wallach und Kogan, Gruppentestung unter sonst gleichen Bedingungen bei Krause) nur durchschnittliche Korrelationen zwischen den Subtestergebnissen von .41 zu erreichen waren, können für den Zusammenhang zwischen 2 zu verschiedenen Zeitpunkten bearbeiteten Tests nur sehr mäßige Koeffizienten erwartet werden, zumal der Test von Wallach und Kogan nur verbal zu beantwortende Aufgaben enthält, die Versuchsform IV, 1 jedoch auch zeichnerisch zu lösende Items. Die im Vergleich zu Subtestergebnissen höhere Reliabilität von Gesamttestresultaten dürfte demgegenüber weniger ins Gewicht fallen.
Aufgrund dieser Überlegungen wurde eine Korrelation zwischen den Ergebnissen der beiden Verfahren unter .4o erwartet; sie sollte aber höher sein als die Korrelation der Versuchsform IV, 1 mit dem in dieser Untersuchung durchgeführten Intelligenztest KLI.

Tabelle 19 zeigt die Korrelationen der beiden Kreativitätstests untereinander und mit einigen ausgewählten Variablen.

---

1) S. dazu S. 92

2) Die durchschnittliche Korrelation der Aufgaben der Versuchsform IV, 1 in einer Gesamtschulstichprobe mit 444 Schülern beträgt bei Berücksichtigung der Koeffizienten, die sich für Produktivität und Originalität derselben Aufgaben ergeben, .33 (vgl. S. 155-156).

Tabelle 19: Korrelationen der Versuchsform IV, 1 mit dem Kreativitätstest von Wallach und Kogan (W und K) und beider Verfahren mit einigen ausgewählten Variablen der Untersuchung in der Bad Dürkheimer Schule (N = 107)

	IV, 1 (Prod.)	IV, 1 (Orig.)	W u. K (Prod.)	W u. K (Orig.)	KLI Teil L	KLI Teil I	Deutschnote	ZR 4+
IV, 1 (Prod.)	-	.786ss*	.335ss	.271ss	.172	.285ss	-.403ss	-.208s
IV, 1 (Orig.)		-	.222s	.148	.092	.183	-.343ss	-.226s
W u. K (Prod.)			-	.895ss	.160	.231s	-.224s	-.225s
W u. K (Orig.)				-	.130	.223s	-.192	-.188
KLI Teil L					-	.590ss	-.276ss	-.268ss
KLI Teil I						-	-.416ss	-.497ss
Deutschnote							-	.462ss
Geschlecht (1=männl. 2=weibl.**)	.113	.151	.234s	.185	.158	-.060	-.166	.043
Vaterberuf (Stufen 1-6 nach Kleining u.Moore 1968;1=Oberschicht)	-.005	.103	.111	.117	.089	-.044	.125	.038

*) Siginifikante Korrelationskoeffizienten ($p \leq .05$) sind durch s, sehr signifikante ($p \leq .01$) durch ss gekennzeichnet.

**) S. dazu Anmerkung 1, S. 172

Die Korrelation von .335 zwischen den Produktivitätswerten der beiden Verfahren entspricht den Erwartungen, nicht jedoch die Korrelation von .148 zwischen den beiden Originalitätswerten. Daß die Zusammenhänge zwischen den Originalitätsscores und allen anderen Variablen

fast durchweg niedriger sind als die Korrelationen mit den Produktivitätswerten, weist vermutlich darauf hin, daß die Originalitätsauswertung weniger reliabel ist als diejenige nach der Anzahl der Lösungen. Aus den Korrelationen zwischen Produktivität und Originalität der beiden Tests kann geschlossen werden, daß dies für Versuchsform IV, 1 in stärkerem Ausmaß zutrifft als für das Verfahren von Wallach und Kogan (s. dazu S. 183).

Beide Verfahren korrelieren untereinander höher als mit den Ergebnissen des Intelligenztests. Die Zusammenhänge mit Teil L des KLI sind nicht signifikant, mit dem Teil I korreliert nur der Produktivitätswert der Versuchsform IV, 1 hochsignifikant, während die Beziehung für den Originalitätswert die Signifikanzgrenze knapp verfehlt; die Korrelationskoeffizienten für den Zusammenhang der Ergebnisse des Tests von Wallach und Kogan mit Teil I des KLI liegen dazwischen und sind beide mit einer Zufallswahrscheinlichkeit von .o5 signifikant.
Die Korrelation der Produktivitätswerte der beiden Kreativitätstests ist zwar höher als diejenige zwischen den Ergebnissen der Versuchsform IV, 1 und Teil I des KLI, aber die Differenz zwischen den beiden Koeffizienten ist nicht signifikant.[1] Auch die Differenz zwischen dem Korrelationskoeffizienten für die Produktivitätswerte der beiden Kreativitätstests und demjenigen für die Beziehung zwischen dem Anzahlwert des Verfahrens von Wallach und Kogan und Teil I des KLI ist nicht signifikant.[2]
Eine Abgrenzung der Ergebnisse der Versuchsform IV, 1 und des Tests von Wallach und Kogan von denjenigen des Intelligenztests im Sinne einer konvergenten und diskriminanten Validierung ist also nicht in dem erwarteten Ausmaß gegeben.

Tabelle 19 zeigt außerdem, daß die beiden Schulleistungsvariablen mit dem KLI, Teil I, enger zusammenhängen als mit den beiden Kreativitätstests und daß die Versuchsform IV, 1 eine engere Beziehung zu diesen Variablen hat als der Test von Wallach und Kogan.

Schließlich wurden noch die Korrelationen sämtlicher Verfahren mit dem Geschlecht der Schüler[3] und dem Beruf des Vaters in die Tabelle aufgenommen. Die Produktivitätsauswertung des Tests von Wallach und

---

1) Überprüfung mit dem t-Test für abhängige Stichproben von Hotelling (1940, beschrieben in Edwards 1971, S. 97-98; t = .448 bei 104 Freiheitsgraden)

2) t = 1.009 bei 104 Freiheitsgraden

3) S. dazu Anmerkung 1, S. 172

Kogan bevorzugt die Mädchen; dieses Ergebnis ist auf dem 5 %-Niveau gesichert. Alle anderen Korrelationen sind nicht signifikant.

Die wichtigsten Ergebnisse dieses Versuchs einer inneren Validierung der Versuchsform IV, 1, der zugleich einen Beitrag zur Konstruktvalidierung lieferte, sind die hochsignifikante, aber nur mäßig hohe Korrelation mit einem als valide geltenden Kreativitätstest und die nicht signifikant niedrigere Korrelation des Verfahrens mit einem Teilergebnis des Intelligenztests KLI.

3.223 Folgerungen aus den Untersuchungsergebnissen

Ziel aller bisher dargestellten Untersuchungen war, ein Verfahren zur Erfassung der Fähigkeit zu divergentem Denken als einer von mehreren Voraussetzungen für kreatives Verhalten zu entwickeln. Die Arbeiten mit der Versuchsform IV, 1 enthalten außerdem erste Ansätze zur Validitätsprüfung.
Die Konsequenzen der Untersuchungsergebnisse für eine erste Beurteilung der Validität sowie für eine Revision des Verfahrens werden im folgenden besprochen.

3.2231 Überlegungen zu den Validitätsuntersuchungen

Mit der Versuchsform IV, 1 wurden sowohl Untersuchungen zur Konstrukt- als auch zur Übereinstimmungsvalidität durchgeführt. Erstere werden als die weitaus wichtigeren für die Validierung eines Kreativitätstests betrachtet, obwohl die Grundlage dieser Konstruktvalidierung - entsprechend dem unbefriedigenden theoretischen Stand der Kreativitätsforschung - nur ein sehr dürftiges "monologisches Netz" ist.[1)]

Das Verfahren, dessen Resultate sicher stärker von motivationalen Bedingungen abhängen als Intelligenztestergebnisse, soll eine möglichst große Zahl von Begabungs- und/oder Interessenrichtungen ansprechen, es muß also viele verschiedenartige Aufgaben enthalten. Trotzdem muß der Test, der eine von mehreren Voraussetzungen für kreatives Verhalten untersuchen soll, ein gewisses Maß an innerer Konsistenz besitzen; nur dann kann angenommen werden, daß er ein einheitliches Konstrukt mißt. Um die Möglichkeit auszuschließen, daß die Aufgaben

---

1) Nach Cronbach und Meehl (1955) ist eine Konstruktvalidierung auch in einem sehr frühen Stadium der Theorieentwicklung möglich: "The logic of construct validation is involved whether the construct is highly systematized or loose, used in ramified theory or a few simple propositions, used in absolute propositions or in probability statements." (S. 284)

nur aufgrund ihres Zusammenhangs mit der Intelligenz der Schüler korrelieren, müssen darüber hinaus die Interkorrelationen der Items der Versuchsform IV, 1 höher sein als ihre Korrelationen mit den Ergebnissen eines Intelligenztests.

Die Ergebnisse der Untersuchung in 15 Gesamtschulklassen zeigen, daß die Versuchsform IV, 1 diese Forderungen im wesentlichen erfüllt. Bei einigen Aufgaben erscheinen allerdings Verbesserungen angezeigt, die unter 3.2232 besprochen werden.
Die verschiedenen Untersuchungen an der Hauptschule in Bad Dürkheim stellen weitere Versuche der Validierung der Versuchsform IV, 1 dar. Sie bestehen aus Überprüfungen der Übereinstimmungsvalidität; eine dieser Überprüfungen kann zugleich als Untersuchung zur Konstruktvalidität betrachtet werden.
Die beiden Ansätze zur äußeren Validierung - der Einsatz eines Lehrerfragebogens zum Unterrichtsverhalten der Schüler und eine Befragung der Schüler über das Verhalten der Mitschüler in 3 kreativitätsfordernden Situationen - erbrachten keine überzeugenden Belege für die Validität der Versuchsform IV, 1 . Das Validitätskriterium Lehrerurteil erwies sich in dieser Stichprobe als ungeeignet. Da die Fragebogen jedoch nur von 4 Lehrern derselben Schule beantwortet worden waren, sollte er in einer weiteren Untersuchung an einer größeren Zahl von Klassen aus mehreren Schulen noch einmal erprobt werden. Um eine sorgfältige Bearbeitung zu sichern, wurde eine nach den Ergebnissen der Faktorenanalyse gekürzte Version des Fragebogens verwendet, nach der jeder Lehrer alle Schüler auf 9 Skalen einzustufen hatte (obwohl nicht auszuschließen war, daß die Faktorenstruktur der ursprünglichen Version des Fragebogens in einer neuen Stichprobe anders ausgefallen wäre als in den Bad Dürkheimer Klassen). Bei den 4 Ergebnissen der Schülerbefragung waren es weniger die absoluten Höhen der Validitätskoeffizienten, die zumindest für die Produktivitätswerte z.T. recht gut, z.T. noch akzeptabel erscheinen, sondern die ebenfalls untersuchten Beziehungen der Außenkriterien zu anderen Variablen, die Zweifel an der Validität der Kriterien aufkommen ließen. Die niedrigsten Korrelationen mit Schulleistungs- und Intelligenzmaßen hatte die Frage 3, die vermutlich wegen einer stark durch das Geschlecht der Schüler beeinflußten Beantwortung den niedrigsten Validitätskoeffizienten hatte. Es sollte nun versucht werden, durch geringfügige Formulierungsänderungen - die Worte

"lustige Streiche" wurden z.B. in der Instruktion des Versuchsleiters durch "lustige Überraschungen" ersetzt - die Geschlechtsabhängigkeit der Schülerratings zumindest zu reduzieren. Es ist anzunehmen, daß sich dadurch die Validität dieses Außenkriteriums erhöht.
Abgesehen von diesen relativ unbedeutenden Veränderungen sollte dieser Teil der Validitätsüberprüfung an einer größeren Stichprobe mit Klassen aus verschiedener Schulen wiederholt werden, da in einer Stichprobe aus 4 fünften Hauptschulklassen eine Streuungseinschränkung im Bereich des divergenten Denkens nicht unwahrscheinlich ist.[1] Gerade in dem z.T. in Hauptschulen überproportional vertretenen unteren und mittleren IQ-Bereich sind auch in ihrer Streuung eingeschränkte (überwiegend niedrige bis mittelmäßige) Kreativitätstestergebnisse zu erwarten (vgl. dazu S. 8, S. 73-74 und 93- 95). In einer Stichprobe mit stärkerer Streuung der Kreativitätstestergebnisse müßten die Validitätskoeffizienten höher ausfallen.

Das Argument einer eventuellen Streuungseinschränkung bei den Kreativitätstestergebnissen läßt es auch angezeigt erscheinen, den Test von Wallach und Kogan in einer größeren Stichprobe noch einmal durchzuführen. Es wird allerdings trotzdem ein Validitätskoeffizient unter .4o erwartet (vgl. dazu S. 187-188).
Eine kritische Beurteilung der Ergebnisse der unter 3.22 dargestellten Validierungsuntersuchungen läßt es angezeigt erscheinen, für die Validitätsprüfung der verbesserten Version des Tests zusätzliche und von den bisher verwendeten möglichst verschiedene Kriterien einzusetzen.

3.2232 Revision der Versuchsform IV, 1

Trotz der insgesamt zufriedenstellenden Ergebnisse der ersten Validitätsprüfungen besteht wahrscheinlich die Möglichkeit, durch Revision einiger Aufgaben ihre Trennschärfe und ihre Reliabilität zu erhöhen und auf diese Weise die Validität des Verfahrens zu verbessern.

---

1) Ebenso kann angenommen werden, daß die Streuung der Intelligenz in dieser Stichprobe eingeschränkt ist, da ein großer Teil der intelligenteren Schüler nach der 4. Klasse in Gymnasien oder Realschulen überwechselt. Eine Überprüfung dieser Annahme war nicht möglich, da die Eichung des KLI in 4. Klassen und in 5. Hauptschulklassen durchgeführt wurde, also ebenfalls in einer Stichprobe mit reduzierter Streuung der Intelligenz.

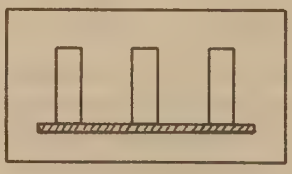
Nach den Ergebnissen der Korrelationsberechnungen und der Faktorenanalyse ist in erster Linie die Aufgabe 1 revisionsbedürftig.[1)] In der Untersuchung der Trennschärfe der Items mit dem Programm DATANA ergaben sich für die Produktivitätswerte der Aufgaben 1, 2 und 4 die niedrigsten Koeffizienten.[2)] (Die Trennschärfe der Originalitätsscores liegt immer unter derjenigen des Produktivitätswertes derselben Aufgabe.)

Eine weitere Entscheidungshilfe bot die Inspektion der von DATANA ebenfalls ausgedruckten Verteilungen der Testergebnisse. Während die Originalitätswerte zum großen Teil stark linksgipflige Verteilungen zeigten, traf dies bei den Produktivitätsscores in erster Linie für die Aufgaben 1, 2, 4 und 5 zu; der Modalwert dieser Verteilungen lag bei 2 (bei Antwortkategorien von o bis 8 [3)]).

Aufgrund dieser Ergebnisse wurden die Aufgaben 1, 2 , 4 und 5 z.T. modifiziert, z.T. durch andere Items ersetzt.

Für eine Verbesserung der Aufgabe 1 schienen die nicht akzeptablen Lösungen einen Anhaltspunkt zu bieten. Ein Teil der Schüler hatte die in der Instruktion enthaltene Anweisung "Wenn man baut, soll nichts umfallen. Zeichne Deine Figuren auch so, daß die Klötze stehen bleiben" nicht beachtet und Figuren gezeichnet, die nicht stehen bleiben. Es wird vermutet, daß diese Kinder während des Zeichnens von der Vorstellung ausgingen, von oben auf die Figuren zu sehen, mit anderen Worten, mit liegenden Klötzen zu bauen. Um ein solches Aufgabenverständnis auszuschließen, wurde die Zeichnung der Klötze etwas modifiziert und kurz erklärt:

Du siehst, die Klötze sind länglich und stehen auf einem Brett.



Aufgabe 2, die bei vielen Kindern ziemlich angsterregende Vorstellungen erzeugte, wurde durch eine andere ersetzt, in der sich die Probanden

---

1) S. Tabelle 8, S. 157 (Die Ergebnisse der Korrelationsanalysen wurden wegen ihrer Unübersichtlichkeit nicht in ihrer ursprünglichen Form mitgeteilt.)

2) Vgl. S. 162

3) S. dazu S. 16o-161

alle möglichen Folgen ausdenken sollen, die die Erfindung von Tabletten, die alle anderen Nahrungsmittel überflüssig machen, haben könnte. Die Reihenfolge der beiden Aufgaben wurde vertauscht, da die Aufgabe 3 der Versuchsform IV, 1 einen leichteren Zugang zu diesem Aufgabentyp zu bieten schien als die Tablettenaufgabe.[1)]

Aufgabe 4, deren Auswertung die größten Schwierigkeiten bereitet hatte, wurde gestrichen. Da eine weitere Verkürzung des Tests nicht vertretbar erschien, wurde statt der Aufgabe 4 eine 3. Aufgabe vom Typ "Bilderraten" aufgenommen. Diese neue Aufgabe 11 [1)] wurde aus dem Kreativitätstest von Wallach und Kogan übernommen (Aufgabe 6 des Untertests "line meanings"), da sie im Hinblick auf Unabhängigkeit vom Geschlecht und vom Vaterberuf, auf Trennschärfe (nach DATANA) und Höhe der Korrelation mit dem KLI, Teil I (für Produktivität r =.13, für Originalität r =.06) in der Untersuchung in der Bad Dürkheimer Schule günstig abgeschnitten hatte.
Die Aufgabe 5 wurde ebenfalls durch eine von Wallach und Kogan übernommene Aufgabe ersetzt (Aufgabe 2 des Subtests "similarities")[1)]. Die Auswahlkriterien waren dieselben wie bei Aufgabe 11, allerdings korreliert die Originalitätsauswertung dieser Aufgabe signifikant mit dem Geschlecht (r = .21, d.h. die Mädchen gaben originellere Antworten als die Jungen).

Die Instruktion zu den Items 5 und 6 der Versuchsform IV, 1 wurde so umformuliert, daß die abstrakte Erläuterung "es kommt darauf an, möglichst viele Gründe für die Vergleiche zu finden" wegfallen konnte. Außerdem wurde eine andere Beispielaufgabe gewählt, unter deren Lösungen auch einige ausgefallenere sind.
Die beiden Aufgaben werden in der neuen Testform, die im folgenden als Versuchsform VI, K bezeichnet wird, an 7. und 8. Stelle bearbeitet, da nach Wegfall der Aufgabe 4 die Aufgabengruppe "Zeichnungen" zwischen die beiden verbalen Aufgabengruppen "Verrückte Geschichten" und "Lustige Vergleiche" placiert wurde.[1)]
Auch bei der Aufgabengruppe "Zeichnungen" wurde versucht, die Instruktion etwas zu vereinfachen[2)], die Aufgaben selbst blieben unverändert.

---

1) S. Anhang A 6

2) Ein sinnentstellender Tippfehler in der Instruktion der Versuchsform IV, 1 ("darauf"statt "daraus") scheint das Aufgabenverständnis der Kinder nicht oder wenig beeinträchtigt zu haben.

Aus den Aufgaben 1o und 11 der Versuchsform IV, 1 wurden ohne Änderung die Aufgaben 9 und 1o der Versuchsform VI, K; an sie schließt sich die bereits erwähnte Aufgabe 11 an, die aus dem Kreativitätstest von Wallach und Kogan übernommen wurde.

3.3 Testvalidierung

Die Darstellung der mit der Versuchsform VI, K durchgeführten Untersuchungen konzentriert sich auf die verschiedenen Überprüfungen der Validität als dem wichtigsten und zugleich problematischsten Testgütekriterium.
Die Reliabilität im Sinn der internen Konsistenz wird zwar bei der Untersuchung der diskriminanten Validität erneut überprüft, aber nicht mehr ausführlich besprochen, da dieses Gütekriterium aufgrund der Untersuchungsergebnisse der Versuchsform IV, 1 ausreichend gesichert erscheint.
Die Auswertungsobjektivität der Versuchsform VI, K wurde ebenfalls untersucht und ist in Anbetracht der nicht völlig objektivierbaren Auswertung als gut zu bezeichnen.[1)]
Nach einem Überblick über die mit der Versuchsform VI, K durchgeführten Untersuchungen werden die damit verfolgten Untersuchungsziele besprochen. Dabei werden auch Untersuchungshypothesen aufgestellt, die sich allerdings im Hinblick auf die Bestimmtheit der Formulierung von den Hypothesen experimenteller Forschungsarbeiten unterscheiden. Aus diesem Grund wurde auch auf die Formulierung der den geäußerten Erwartungen entsprechenden Nullhypothesen verzichtet.

3.31 Überblick über die Untersuchungen

Die Validierung des Tests zur Erfassung des divergenten Denkens gliedert sich in 2 Teile: in die
Untersuchungen mit der Gesamtstichprobe (N=1671) und in die
Untersuchungen mit einer Teilstichprobe aus 5 Gesamtschul-Klassen (N = 137).
Die Gesamtstichprobe bestand aus Schülern fünfter Klassen aus
Gesamtschulen (N_G = 873),
Realschulen (N_R = 4o8),
Gymnasien (N_{Gym} = 296),
Hauptschulen (N_H = 94).
Zur Validierung der Versuchsform VI, K wurden in dieser Stichprobe folgende Verfahren eingesetzt:

1) Diese Untersuchungsergebnisse werden in Anhang B 1 dargestellt.

Das Prüfsystem für Schul- und Bildungsberatung P-S-B von Horn (1969) und der im Zentrum für empirische pädagogische Forschung entwickelte Bildungs-Beratungs-Test BBT. Außerdem wurden von den für alle Schüler erhobenen Informationen folgende berücksichtigt:
Alter (in Jahren mit 2 Dezimalstellen),
Geschlecht,
Angstfragebogen für Schüler (AFS) von Wieczerkowski, W., Nickel, H., Janowski, A., Fittkau, B. und Rauer, W.,
Beruf des Vaters,
Schulbildung des Vaters,
Schulbildung der Mutter,
sozioökonomischer Status (ein aus den drei zuletzt genannten Variablen berechnetes Status-Maß; sämtliche 4 Angaben zur sozioökonomischen Situation der Schüler wurden nach einem von Bauer 1971 dargestellten Verfahren erhoben),
die Zensuren in Deutsch, Mathematik, erster Fremdsprache und Kunsterziehung vom Halbjahres- und Jahreszeugnis der 5. und vom Halbjahreszeugnis der 6. Klasse.

Die Teilstichprobe bestand aus 3 fünften Klassen hessischer Gesamtschulen und aus 2 fünften Gesamtschulklassen aus Baden-Württemberg.[1] In diesen 5 Klassen wurden - zusätzlich zu den in der Gesamtstichprobe durchgeführten - folgende Verfahren angewandt: Ein Lehrerfragebogen zum Unterrichtsverhalten der Schüler,[2] eine Schülerbefragung über das Verhalten aller Schüler in 3 kreativitätsfordernden Situationen,[3] der Kreativitätstest von Wallach und Kogan[4] und der Persönlichkeitsfragebogen CPQ von Porter und Cattell (deutsche Übersetzung 1968). In den folgenden Abschnitten werden die verwendeten Verfahren - soweit dies nicht schon unter 3.222 geschehen ist - erläutert und ihr Einsatz in dieser Validitätsuntersuchung begründet.

3.32 Untersuchungsziele

Die Untersuchungen in beiden Stichproben dienten in erster Linie der Validitätsprüfung des Verfahrens.

1) Das gesamte Untersuchungsmaterial war an 8 Klassen, die ihre Mitarbeit zugesagt hatten, verschickt worden. Von einer dieser Klassen erhielten wir trotz wiederholter Bitten nur den Kreativitätstest von Wallach und Kogan ausgefüllt zurück, in einer weiteren Klasse - der einzigen von insgesamt 57 - waren die Aufgaben 4 - 6 falsch bearbeitet worden und eine dritte Klasse war für alle angewandten Verfahren so schlecht motiviert (die Aufgaben wurden z.T. überhaupt nicht bearbeitet und statt dessen mit aggressiven und/oder obszönen Randbemerkungen versehen), daß es angezeigt erschien, die Ergebnisse nicht zu verwenden.

2) S. dazu unter 3.2221, S. 165-171

3) S. dazu unter 3.2222, S. 178-181

4) S. dazu unter 3.2223, S. 187-188

3.321 Untersuchungen in der Gesamtstichprobe

Die Untersuchungen in der Gesamtstichprobe sind z.T. als direkte Validitätsprüfungen zu betrachten, z.T. stellen sie eher Überprüfungen der Voraussetzungen für die Gültigkeit des Tests zur Erfassung des divergenten Denkens dar.

3.3211 Überprüfung der diskriminanten Validität des Verfahrens

Für die von der Versuchsform IV, 1 in mehreren Aufgaben abweichende Versuchsform VI, K mußte nachgewiesen werden, daß die Items ein einheitliches Konstrukt messen, das sich deutlich von den durch Intelligenztests gemessenen Persönlichkeitsdimensionen unterscheidet.
Dies sollte wie bei Versuchsform IV, 1 durch einen Vergleich der durchschnittlichen Korrelationen der Kreativitätstest- und Intelligenztestaufgaben und durch eine Faktorenanalyse geschehen.[1] Da die am wenigsten trennscharfen Items der Versuchsform IV, 1 modifiziert (Aufgabe 1) bzw. ausgeschieden worden waren (Aufgabe 2 und 4), war anzunehmen, daß die interne Konsistenz der Versuchsform VI, K höher ist als diejenige der Versuchsform IV, 1.
Für den Zusammenhang zwischen den Kreativitätstestitems und den beiden Intelligenztests wurden für die Gesamtstichprobe höhere Koeffizienten erwartet als in dem aus 17 Gesamtschulklassen bestehenden Sample, da die neue Stichprobe neben Gesamtschul- auch Hauptschul-, Realschul- und Gymnasialklassen enthält, was sehr wahrscheinlich eine Erhöhung der Testwertstreuung zur Folge hat. Die gemittelten Korrelationen der Aufgaben der Versuchsform VI, K mit den Subtests der beiden Intelligenztests sollten aber niedriger sein als der Mittelwert der Interkorrelationen der VI,K-Items untereinander.
Bezüglich der Ergebnisse der Faktorenanalyse wurde erwartet, daß die Kreativitätsaufgaben auf anderen Faktoren am höchsten laden als die Intelligenztest- und Schulleistungsvariablen und daß sich die klare Trennung zwischen den Items der Versuchsform VI, K und allen übrigen Variablen auch bei der modifizierten Aufgabe 1 zeigt.

3.3212 Überprüfung der Geschlechtsabhängigkeit des Verfahrens

Anhand der großen und alle Schultypen umfassenden Stichprobe sollte überprüft werden, ob die Kreativitätstestaufgaben Jungen oder Mädchen

---

1) Eine nochmalige Berechnung der Reliabilitäts- und Trennschärfe-Koeffizienten mit dem Programm DATANA war wegen der dafür erforderlichen Transformation der Ausgangsdaten (s. dazu S. 160-161) nicht möglich.

begünstigen. Eine erste Überprüfung dieser Frage war in der Untersuchung an der Bad Dürkheimer Schule erfolgt und hatte ein negatives Ergebnis erbracht. Da jedoch die Bad Dürkheimer Stichprobe sehr klein war und nur aus Hauptschülern bestand und weil einige Testaufgaben modifiziert bzw. durch andere ersetzt wurden, erschien eine zweite Überprüfung dieser Frage erforderlich.

Falls sich zeigen sollte, daß die Versuchsform VI, K eines der beiden Geschlechter begünstigt, so wurde ein besseres Abschneiden der Mädchen erwartet, da in die neue Versuchsform 2 Aufgaben aus dem Kreativitätstest von Wallach und Kogan übernommen worden waren, von denen in der Bad Dürkheimer Untersuchung eine - die neue Aufgabe 5 - ebenso wie der Gesamtwert für Produktivität signifikant mit dem Geschlecht korrelierten (s. Tabelle 19, S. 189).
Das bessere Abschneiden der Mädchen, das im übrigen auch in einigen anderen Untersuchungen[1] festgestellt wurde, müßte als Hinweis darauf betrachtet werden, daß das Verfahren neben der Fähigkeit zu divergentem Denken noch eine oder mehrere Variablen erfaßt, in denen Mädchen höhere Werte erreichen (wie z.B. schulische Anstrengungsbereitschaft).

3.3213 Überprüfung der Schichtabhängigkeit des Verfahrens

Obwohl in der Bad Dürkheimer Stichprobe weder der Kreativitätstest von Wallach und Kogan noch die Versuchsform IV, 1 mit dem in Anlehnung an Kleining und Moore (1968) entwickelten Maß des sozioökonomischen Status korrelierten und obwohl dieses Ergebnis von Krause (1972) für den Test von Wallach und Kogan und ein anderes Status-Maß bestätigt wird,[2] soll an dieser großen und im Hinblick auf die sozioökonomische Position der Eltern weit gestreuten Stichprobe überprüft werden, ob die Versuchsform VI, K einen schichtspezifischen Bias hat. Um auszuschließen, daß ein negatives Ergebnis nur deshalb zustande kam, weil der verwendete Statusindex zu wenig mit dem für kreative Fähigkeiten bedeutsamen Erzieherverhalten zusammenhängt, wurden 3 soziodemographische Einzeldaten erhoben - der Beruf des Vaters und die Schulabschlüsse beider Eltern - und ein auf diesen Einzeldaten basierendes kombiniertes Statusmaß berechnet - der "Index des für Bildungsverhalten relevanten sozialen Status" ("Bauer-Status", vgl. Bauer 1971).

---

1) S. dazu S. 66-67 und S. 136-137

2) Vgl. Krause 1972, S. 119

3.3214 Überprüfung der diagnostischen Valenz der Originalitätswerte

Für die weitere Arbeit mit dem Kreativitätstest stellt sich schließlich noch eine Frage von erheblicher praktischer Bedeutung: Bringen die mit erheblichem Arbeitsaufwand ermittelten Originalitätswerte[1] reliable Informationen, die in den Produktivitätswerten nicht enthalten sind? Sämtliche Originalitätsscores der Versuchsform IV, 1 waren weniger trennscharf als die zugehörigen Produktivitätswerte (vgl. Tabelle 1o, S. 162), was eventuell allein auf die geringere Auswerterobjektivität bei der Vergabe der Originalitätspunkte zurückzuführen ist.[2]
In den Validitätsuntersuchungen ergaben sich für die Produktivitätswerte teils höhere Korrelationskoeffizienten, teils bewegten sich die Koeffizienten der Produktivitäts- und Originalitätsscores in gleicher Höhe (vgl. Tabelle 13, S.172 , Tabelle 17, S. 182 , Tabelle 19, S.189). Es ergaben sich keine Hinweise darauf, daß durch die Originalitätswerte ein anderer Aspekt kreativer Fähigkeiten erfaßt wird als durch die Produktivitätswerte.

Für die Ergebnisse der Versuchsform VI, K wurde ein engerer Zusammenhang zwischen Anzahl- und Originalitätswerten erwartet als für diejenigen der Versuchsform IV, 1. Einmal führt die stärkere Testwertstreuung der 4 Schultypen umfassenden Stichprobe sehr wahrscheinlich zu engeren Beziehungen zwischen den meisten Untersuchungsergebnissen, zum anderen ist anzunehmen, daß einige inzwischen vorgenommene kleinere Präzisierungen und Ergänzungen der Auswertungsrichtlinien eine Erhöhung der Produktivitäts-Originalitäts-Korrelation zur Folge haben.
Das Hauptziel der Ergänzungen war, zu verhindern, daß durch geringfügige Abwandlungen eines einzigen Einfalls relativ hohe Produktivitätsscores erreicht werden können.[3] Deshalb wurden die Auswerter angewiesen, nur bis zu 2 solcher Abwandlungen einer Antwort zu werten.[4]

---

1) S. dazu S. 152, Anmerkung 1

2) Natürlich ist auch die Produktivitätsauswertung nicht völlig objektiv, aber die Vergabe der Originalitätspunkte baut auf dieser auf und bringt zusätzliche Fehler mit sich.

3) Beispiele für solche Abwandlungen derselben Lösung sind: bei Aufgabe 4 "Teller mit Brei,Teller mit Suppe, Teller mit Gemüse,..."; bei Aufgabe 8 "Herz, Magen, Lunge, Leber, Galle,...".

4) S. Anhang A 7

Es wurde angenommen, daß die Berücksichtigung dieses minimalen Qualitätskriteriums zu einer leichten Erhöhung des Zusammenhangs zwischen Anzahl- und Originalitätswerten führt, da auf diese Weise verhindert wird, daß insgesamt wenig produktive Kinder, die nur bei einer Aufgabe oder jedenfalls bei sehr wenigen Aufgaben durch weitgehende Ausschöpfung einer Lösungskategorie auf statistisch seltene Antworten kommen, mehr Originalitätspunkte erhalten als nach der Gesamtzahl ihrer Einfälle zu erwarten ist.
Die Entscheidung über Beibehaltung oder Wegfall der Originalitätsauswertung sollte von der Höhe der Korrelation mit den Produktivitätswerten abhängig gemacht werden. Falls die Korrelation der beiden Gesamtwerte für die Versuchsform VI, K in der großen Stichprobe .90 erreichen sollte, erschien es überflüssig, die Antworten der Kinder weiterhin nach beiden Kriterien auszuwerten; die Originalitäts-Auswertung könnte dann entfallen. Bei Korrelationen der beiden Werte zwischen .80 und .90 sollte noch einmal überprüft werden, ob die Originalitätsauswertung wirklich keine zusätzlichen Informationen erbringt; nur wenn diese Frage verneint werden kann, sollte der Test in Zukunft nicht mehr im Hinblick auf die Originalität der Antworten ausgewertet werden.

3.322 Untersuchungen in der Teilstichprobe

Die nochmalige Verwendung des Lehrerfragebogens, der Schülerbefragung und des Kreativitätstests von Wallach und Kogan als Validitätskriterien in einer Stichprobe von 5 Klassen aus 3 verschiedenen Schulen wurde unter 3.2231 begründet (s. dazu S.192-193).

3.3221 Überprüfung der Übereinstimmungsvalidität mit dem Lehrerfragebogen

Bezüglich der Ergebnisse des Lehrerfragebogens sollte die Hypothese, daß die Kreativitätstestergebnisse mit dem ersten Gesamtwert des Lehrerfragebogens ("Motivation für kreatives Verhalten") positiv und substantiell korrelieren, noch einmal überprüft werden. Für den zweiten Gesamtwert der Lehrerratings wurde lediglich erwartet, daß die Beziehungen zu den Kreativitätstestergebnissen weniger eng sind. Die abweichenden Ergebnisse in der Bad Dürkheimer Schule sind vermutlich auf Besonderheiten in der Beurteilungsweise der 4 Lehrer oder auf schulspezifische Urteilsschemata zurückzuführen.

3.3222 Überprüfung der Übereinstimmungsvalidität mit der Schülerbefragung

Für die 4 Ergebnisse der Schülerbefragung wurden zumindest ebenso hohe Korrelationen mit den Kreativitätstestergebnissen erwartet wie in der Bad Dürkheimer Untersuchung; die 3. Frage ("lustige Überraschungen ausdenken") sollte nicht mehr so stark die Jungen begünstigen und infolgedessen einen etwas engeren Zusammenhang mit den Ergebnissen der Versuchsform VI, K aufweisen. Maßnahmen, die die Beeinflussung der Ratings durch die Schulleistungen vermindern, erschienen nicht angezeigt, da die Schulleistungen vermutlich auch in Gesamtschulen die weitaus wichtigste Beurteilungsgrundlage sein dürften.

3.3223 Überprüfung der Konstruktvalidität mit dem Test von Wallach und Kogan und zwei Intelligenztests

Ein unerwartetes Ergebnis der Bad Dürkheimer Untersuchung war die niedrige Korrelation zwischen den Ergebnissen der Versuchsform IV, 1 und denjenigen des Kreativitätstests von Wallach und Kogan (vgl. Tabelle 19, S.189). Eine mögliche Erklärung dafür schien in der eingeschränkten Streuung der kreativen Fähigkeiten in dieser Stichprobe zu liegen (s. dazu S. 193). Da die Streuungseinschränkung in 5 Gesamtschulklassen sicher geringer ist als in 4 Hauptschulklassen derselben Klassenstufe und nachdem 2 Items aus dem Kreativitätstest von Wallach und Kogan in die Versuchsform VI, K aufgenommen worden waren, wurde für die Hauptuntersuchung eine Korrelation von $r > .45$ zwischen den Produktivitätswerten der beiden Verfahren erwartet. Falls eine weitere Berücksichtigung der Originalitätswerte erforderlich sein sollte (s. dazu S. 2oo-2o1), war wegen der geringeren Objektivität der Originalitätsauswertung anzunehmen, daß der Zusammenhang zwischen den Originalitätsscores weniger eng ist, aber der Korrelationskoeffizient müßte auf jeden Fall den entsprechenden Wert der Bad Dürkheimer Untersuchung übertreffen; es wird ein r zwischen .4o und .45 erwartet.

Die Überprüfung der konvergenten Validität sollte ergänzt werden durch eine Untersuchung der diskriminanten Validität: Es wurde eine signifikant höhere Beziehung zwischen den beiden Tests zur Erfassung des divergenten Denkens erwartet als zwischen diesen und den beiden Intelligenztests P-S-B und BBT.

3.3224 Überprüfung der Konstruktvalidität mit dem Persönlichkeitsfragebogen CPQ

Mit den bisher aufgeführten Verfahren wird ein weiter Bereich intellektueller Faktoren untersucht. Mit Hilfe des Lehrerfragebogens wird darüber hinaus versucht, Motivationen für verschiedene schulische Aktivitäten zu erfassen, die Rückschlüsse auf die Motivation für kreatives Verhalten ermöglichen.
Abgesehen von dieser Untersuchung der Motivation der Schüler, deren Erfolg nach den Ergebnissen der Voruntersuchung zweifelhaft war, waren bisher keine Verfahren zur Erfassung nichtintellektueller Persönlichkeitsdimensionen eingesetzt worden.
Da jedoch kreatives Verhalten stärker als z.B. intelligentes Verhalten (im Sinn von überwiegend konvergentem Denken) durch nichtintellektuelle Persönlichkeitsfaktoren beeinflußt wird und da zu diesem Forschungsgebiet bereits eine Reihe von empirischen Ergebnissen vorliegt - vor allem von Cattell und seinen Mitarbeitern - sollte versucht werden, die Konstruktvalidierung des Kreativitätstests durch eine Untersuchung einiger Persönlichkeitsmerkmale zu ergänzen, für die Zusammenhänge mit kreativem Verhalten bereits nachgewiesen wurden.
Für diese Untersuchung sollte in Anbetracht der unbefriedigenden Ergebnisse des Lehrerfragebogens und der Schülerbefragung kein ad hoc entwickeltes und nicht oder ungenügend validiertes Verfahren verwendet werden, sondern es wurde ein bereits vorliegender Persönlichkeitsfragebogen gesucht, der nach theoretischen Erwartungen und empirischen Ergebnissen für die Kreativität bedeutsame Dimensionen untersucht.
Von den wenigen für 1o- bis 12Jährige geeigneten Verfahren schied der HANES, KJ ("Hamburger Neurotizismus- und Extraversionsskala für Kinder und Jugendliche" von F. Buggle und F. Baumgärtel, Göttingen: Hogrefe 1972) aus, weil er zwei Persönlichkeitsdimensionen untersucht, über deren Beziehungen zu kreativem Verhalten nur globale theoretische Aussagen, die nicht mit diesem Verfahren überprüft wurden, vorzuliegen scheinen (s. dazu z.B. Krause 1972, S. 58-61).
Für die Altersgruppe der 1o- bis 12 Jährigen liegen außerdem einige Angstfragebogen vor, die z.T. Schul- und Prüfungsangst,[1] z.T. allgemeine Angstbereitschaft bzw. Ängstlichkeit[2] oder beides[3] messen.

---

1) Fragebogen für Schüler FS von Viola Gärtner-Harnach aus der Reihe "Deutsche Schultests", herausgegeben von K. Ingenkamp, Weinheim: Beltz 1973

2) Der Kinder-Angst-Test K-A-T von F. Thurner und U. Tewes, Göttingen: Hogrefe 1969

3) Angstfragebogen für Schüler AFS von W. Wieczerkowski, H. Nickel, A. Janowski, B. Fittkau und W. Rauer, Braunschweig: Westermann 1974.

Nach den vorliegenden Untersuchungsergebnissen ist nicht zu erwarten, daß eine Untersuchung der Schulangst und/oder der Ängstlichkeit der Schüler zur Validitätsprüfung der Versuchsform VI, K beitragen kann.
Das Ergebnis von Wallach und Kogan (1965), daß die Jungen mit geringer Testangst und hoher Abwehrtendenz die niedrigsten Kreativitätstestwerte haben, wurde von Krause (1972) an einer Stichprobe bundesdeutscher Schüler überprüft; er erhielt weder für die Beziehung von Kreativität zu Schulangst noch für den Zusammenhang zwischen Kreativität und Abwehrneigung signifikante Korrelationen, wohl aber eine hochsignifikante negative Beziehung zwischen Schulangst und Abwehrtendenz. Beide Persönlichkeitsdimensionen sind nach diesen Ergebnissen wegen der Komplexität der Zusammenhänge keine geeigneten Kriteriumsvariablen für die Validitätsprüfung der Versuchsform VI, K.
Das einzige Verfahren, das auf dieser Altersstufe Variablen erfaßt, die nach theoretischen Überlegungen und empirischen Untersuchungsergebnissen direkte Beziehungen zur Kreativität haben, scheint z.Zt. der Persönlichkeitsfragebogen CPQ von Porter und Cattell zu sein. Er ist die für 8- bis 12jährige Kinder geeignete Version des 16 PF (Sixteen Personality Factor Questionnaire) und liegt seit 1968 in einer deutschen Übersetzung vor, aber das Verfahren ist weder adaptiert noch an einer bundesdeutschen Schülerstichprobe validiert worden, noch liegen Normdaten für deutsche Kinder vor.
Da nach Cattell und Butcher (1968) mehrere Skalen des CPQ in Untersuchungen an nordamerikanischen Schülern zwischen Hochbegabten und durchschnittlichen Schülern differenzierten und die Persönlichkeitsstruktur der Hochbegabten-Gruppen derjenigen kreativer Wissenschaftler, Lehrer, Verwaltungsbeamter, Künstler und Schriftsteller sehr ähnlich war, wird das Verfahren trotz der Bedenken wegen seiner mangelnden Erprobung an deutschen Schülern in der vorliegenden Untersuchung verwendet.
Die Hypothesen darüber, welche Faktoren des CPQ mit den Kreativitätstestergebnissen auf welche Weise zusammenhängen, wurden vor allem aus theoretischen Annahmen über Persönlichkeitsmerkmale kreativer Individuen und aus Untersuchungsergebnissen zur Persönlichkeitsstruktur hochbegabter Kinder abgeleitet.
Untersuchungen mit dem CPQ und/oder dem HSPQ (High School Personality Questionnaire, also die ab 13 Jahren angemessene Version des 16 PF) wurden von Porter (1964), Werner (1966) und Werner und Bachtold (1969) durchgeführt. Tabelle III in Anhang B gibt einen tabellarischen Überblick über die Ergebnisse der 3 Untersuchungen.
Obwohl die Kriterien für die Auswahl hochbegabter Kinder in diesen

Untersuchungen wahrscheinlich stark von den Schulleistungen abhängen,[1] ist anzunehmen, daß es sich bei diesen Schülern um kreative Kinder handelt. Diese Annahme gründet sich in erster Linie darauf, daß "the CPQ profile of gifted boys in grades 4 - 5 bore a striking resemblance to the 16 PF profile of creative researchers, artists and writers studied by Cattell and Drevdahl (1955, 1958)". (Werner und Bachtold 1969, S. 177).

Daß für Mädchen eine weniger gute Übereinstimmung der CPQ-Ergebnisse mit den Persönlichkeitsprofilen kreativer Erwachsener gefunden wurde,[2] läßt es angezeigt erscheinen, diesen Teil der Validitätsüberprüfung nur mit den Jungen der Unterstichprobe durchzuführen. Die CPQ-Ergebnisse der Mädchen sollen allerdings ebenfalls mitgeteilt, diskutiert und mit denjenigen der Untersuchungen von Porter, Werner und Werner und Bachtold verglichen werden, weil sie - für den Fall einer erfolgreichen Testvalidierung - erste Ansatzpunkte für die Bildung von Hypothesen über Persönlichkeitscharakteristika kreativer Mädchen bieten könnten.

Die Aufstellung von Hypothesen über die Zusammenhänge zwischen den CPQ-Faktoren und den Ergebnissen der Versuchsform VI, K wird durch mehrere Umstände erschwert:

a) Die Skalen des CPQ sind noch nicht für deutsche Kinder validiert und normiert, ihre Beziehungen zur Kreativität sind noch nicht untersucht. Die ungeprüfte Frage der interkulturellen Generalisierbarkeit der amerikanischen Untersuchungsergebnisse verleiht diesem Teil der Validitätsüberprüfung ein gewisses Maß an Unsicherheit.

b) Cattell und Butcher (1968) stellen zwar auch für Schüler eine Spezifikationsgleichung zur Vorhersage von kreativem Verhalten auf,[3] aber sie gehen dabei von den Faktoren des 16 PF aus, was zur Folge hat, daß die berücksichtigten Persönlichkeitsdimensionen weder mit denjenigen des CPQ noch mit denjenigen der Unter-

---

1) Auswahlkriterium dieser 3 Untersuchungen war die Teilnahme an irgendwelchen Spezialprogrammen für begabte Kinder, die u.a. gute Schulleistungen voraussetzte. Hinzu kamen bei Porter (1964) die Empfehlung des Schulleiters und ein IQ (WISC) ab 130, bei Werner (1966) besondere Leistungen auf einem bestimmten Gebiet und bei Werner und Bachtold (1969) hohe Intelligenztestergebnisse. Trotz dieser überwiegend an Schulleistung und Intelligenz orientierten Stichprobenauswahl wird in allen 3 Untersuchungen ein ausdrücklicher Bezug zur Kreativität hergestellt.

2) Das ist nicht verwunderlich, da Cattell und Drevdahl (1955) und Drevdahl und Cattel (1958) nur männliche Kreative untersuchten.

3) Die Spezifikationsgleichung stellt das Fazit der Untersuchungen Cattells und seiner Mitarbeiter mit Schülern dar; sie gibt Ausmaß und Richtung der Zusammenhänge zwischen Kreativität und 12 der 16 Persönlichkeitsfaktoren des 16 PF an. Creativity

suchungen von Porter (1964), Werner (1966) und Werner und Bachtold (1969) völlig übereinstimmen.

c) Nur die Spezifikationsgleichung von Cattell und Butcher ist direkt auf das hier interessierende Kriterium - kreatives Verhalten - bezogen; die Kriteriengruppen der 3 Untersuchungen wurden aufgrund herausragender, im Rahmen der Schule erbrachter Leistungen ausgewählt, wodurch nicht mit völliger Sicherheit gewährleistet ist, daß es sich um kreative Kinder handelte.

d) Die Spezifikationsgleichung von Cattell und Butcher wird nicht für Jungen und Mädchen getrennt aufgestellt, was aufgrund der Ergebnisse der Untersuchungen von Wallach und Kogan(1965), Kogan und Morgan (1969) u.a. und nach den Resultaten der 3 Untersuchungen unbedingt erforderlich wäre.

e) Die Ergebnisse der 3 Untersuchungen stimmen weder untereinander völlig überein noch bestätigen sie in allen Punkten Vorzeichen und Gewichtung der Faktoren der Spezifikationsgleichung.

f) Die Versuchsform **VI**, K wie auch die anderen bisher vorliegenden "Kreativitätstests" messen nicht Kreativität, sondern bestenfalls die Fähigkeit zu divergentem Denken als einer von mehreren Voraussetzungen für kreatives Verhalten. Nur wenn diese Fähigkeit bei 9- bis 12Jährigen bereits eine gewisse Kovariation mit den für kreative Leistungen bedeutsamen Persönlichkeitsfaktoren aufweist, ist es überhaupt möglich, einen Persönlichkeitsfragebogen im Rahmen der Konstruktvalidierung des Verfahrens einzusetzen. Die von Werner und Bachtold (1969) gefundene Übereinstimmung zwischen den Persönlichkeitsprofilen kreativer Erwachsener und denjenigen von hochbegabten Jungen spricht allerdings für eine solche Kovariation, so daß sich diese letzte Schwierigkeit auf die durchgängige Frage der interkulturellen Übertragbarkeit reduziert.

Um einen Überblick über alle bei der Hypothesenbildung berücksichtigten Gesichtspunkte zu ermöglichen, wurden die Hypothesen zu den CPQ-Faktoren zusammen mit den zugrundeliegenden Informationen in Tabellenform dargestellt (s. Tabelle 2o). Aufgrund der erörterten Schwierigkeiten erschien es angezeigt, nur zu den Faktoren, die einen klaren theoretischen Bezug zu kreativem Verhalten haben, Hypothesen über Zusammenhänge mit den Ergebnissen der Versuchsform VI, K aufzustellen.

Dies scheinen beim CPQ vor allem die Faktoren B, C und E zu sein. Zu den Faktoren F, H, I und J werden zwar ebenfalls Hypothesen formuliert; da die Beziehung dieser Faktoren zur Kreativität bei Kindern weniger klar und eindeutig ist, sollen diese Annahmen bei der Validitätsprüfung jedoch ein geringeres Gewicht erhalten.[1] Eine positive Beziehung der Testergebnisse zu den Faktoren G und Q_3 könnte nicht als Validitätshinweis gewertet werden, obwohl die Untersuchungen von Porter (1964), Werner (1966) und Werner und Bachtold (1969) dies nahezulegen scheinen.[2] Nach der für Kinder sehr schulleistungsrelevanten Bedeutung beider Faktoren erscheint die einfachere Erklärung, daß Kinder mit hohen G- und Q_3-Werten infolge ihrer schulischen Leistungsmotivation, Ausdauer und Selbstdisziplin gute Testergebnisse erzielen, plausibler.

---

1) Die für kreative Erwachsene wichtigen Faktoren L (protensia), M (autia), Q_1 (radicalism) und Q_2 (self-sufficiency) werden durch den CPQ nicht erfaßt.

2) Vgl. Tabelle III, Anhang B 2

Tabelle 2c: Übersicht über die Untersuchungshypothesen zu den CPQ-Faktoren und die bei der Hypothesenbildung berücksichtigten Informationen

Faktor		Hypothese	Begründung der Hypothese		
Bezeichnung	Beschreibung *		Kreativitätstheorie(n)	Spezifikationsgleichung	Untersuchung von
A	Kühl, distanziert kritisch (schizothym) versus warm, teilnehmend aus sich herausgehend (cyclothym). Kinder mit hohen Scores für Faktor A reagieren günstiger auf die Schulsituation.	-	-	-	-
B	Wenig intelligent, konkret denkend, geringere schulische Begabung versus intelligent, abstrakt denkend gute schulische Begabung. Die B-Skala ist ein grobes Maß der allgemeinen Intelligenz.	Zwischen den Ergebnissen der Versuchsform VI, K und den Scores für Faktor B besteht eine positive Beziehung. Es wird erwartet, daß B im Vergleich mit den anderen Faktoren des CPQ relativ hoch mit den Testergebnissen korreliert.	Theoretische Annahmen über die Rolle der Intelligenz beim Zustandekommen kreativer Leistungen, s.dazu S. 85.	+ .6 B**	Porter 1964, Werner 1966, Werner und Bachtold 1969** sowie zahlreiche Untersuchungen mit verschiedenen Intelligenz- und Kreativitätsmaßen, in denen überwiegend signifikante positive Korrelationen zwischen allgemeiner Intelligenz und Kreativität gefunden wurden.
C	Geringe Ich-Stärke, emotionale Stabilität und Frustrationstoleranz versus hohe Ich-Stärke, Ausgeglichenheit und Realismus.	Zwischen den Ergebnissen der Versuchsform VI, K und den Werten für Faktor C besteht eine positive Beziehung. Auch für C wird eine relativ hohe Korrelation mit den Testergebnissen erwartet.	Die mit größerer Ich-Stärke verbundene emotionale Stabilität und Frustrationstoleranz sind wesentliche Voraussetzungen für kreative Leistungen (s.dazu z.B. Crutchfield 1962, 1963, Maslow 1959).	+ .3 C	Werner 1966, Werner und Bachtold 1969, Wallach und Kogan 1965; für Erwachsene außerdem Barron 1969, Cattell und Drevdahl 1955, Gough 1961, Mac Kinnon 1965, 1967; weitere Literaturangaben bei Dellas und Gaier 1970.
D	Phlegmatisch, inaktiv, besonnen versus erregbar, ungeduldig, zu übermäßigen Reaktionen neigend.	-	-	-	-
E	Sanft, gehorsam bis unterwürfig, konform versus aggressiv, dominant, unabhängig, anspruchsvoll. Ein hoher Punktwert bedeutet bei Erwachsenen in erster Linie Dominanz, bei Kindern häufiger aggresives Verhalten, das aber durch geeignete Erziehungsmaßnahmen in konstruktive Formen des Sozialverhaltens übergeführt werden kann.	Zwischen den Ergebnissen der Versuchsform VI, K und den Scores für E besteht eine positive und wahrscheinlich relativ enge Beziehung. Wegen des Einflusses der geschlechtsspezifischen Erziehung, die unabhängig-dominantes Verhalten bei Jungen, nicht jedoch bei Mädchen toleriert, sind letztere in diese Hypothese nicht einbezogen***	Die Konformitäts-Unabhängigkeits-Dimension besitzt nach der Theorie von Crutchfield (1962, 1963) eine enge theoretische Beziehung zu kreativem Verhalten.	+ .3 E	Für Jungen wird die Hypothese durch die Untersuchungsergebnisse von Werner und Bachtold (1969) bestätigt. Für Mädchen ergeben sich bei Porter (1964) und Werner (1966) negative Zusammenhänge zwischen E und Kreativität.

*) Die Charkterisierung der Faktoren orientiert sich in erster Linie an dem CPQ-Manual von Porter und Cattell (1968), da einige Faktoren bei Kindern eine etwas andere Bedeutung haben als bei Erwachsene (vor allem N und E, aber auch F, G, Q_3 und Q_4). Zuerst wird der negative, dann der positive Pol der jeweiligen Dimension beschrieben.

**) Vgl. Tabelle III in Anhang B 2

***) Daß Erziehungseinflüsse zu einer für Jungen und Mädchen unterschiedlichen Konstanz von Persönlichkeitsmerkmalen führen, zeigt z.B. die Untersuchung von Kagan und Moss (1962), nach der Aggressivität bei Jungen eine hohe, bei Mädchen eine geringe zeitliche Stabilität besaß, während z.B. die bei beiden Geschlechtern gleichermaßen sozial erwünschte Leistungsmotivation bei Jungen und Mädchen zeitlich stabil war.

Tabelle 2o (Fortsetzung)

Faktor		Hypothese	Begründung der Hypothese		
Bezeichnung	Beschreibung		Kreativitätstheorie(n)	Spezifikationsgleichung	Untersuchung von
F	Nüchtern, ernst, schweigsam, vorsichtig versus enthusiastisch, fröhlich, lebhaft, impulsiv, unbekümmert. Kinder mit hohen F-Werten kommen aus Familien, die ihnen Wärme und Sicherheit und damit Optimismus und Selbstvertrauen geben, während die Kinder mit niedrigen F-Werten unter einem Mangel an Zärtlichkeit und Geborgenheit leiden und infolgedessen ernst und selbstkritisch sind.	Zwischen den Ergebnissen der Versuchsform VI, K und den F-Werten besteht eine positive Beziehung.	Ein Sicherheit vermittelndes, warmes Erziehungsklima stellt eine günstige Voraussetzung für die Entwicklung von Unabhängigkeit und nichtkonformem Verhalten dar, die nach Crutchfield die wichtigsten Bedingungen für kreative Leistungen sind.	Der durch die negative Gewichtung von F gegebene Widerspruch zu der Hypothese und ihren übrigen Begründungen ist eventuell dadurch zu erklären, daß die Untersuchungen Cattells u. seiner Mitarbeiter vor allem mit älteren Stichproben und überwiegend mit Männern durchgeführt wurden. Da F nach Cattell und Butcher (1968) ein stark umweltdeterminierter Faktor ist, kann der Widerspruch eventuell durch die Reaktionen der sozialen Umwelt auf das Verhalten kreativer Kinder und Jugendlicher erklärt werden, die im Lauf der Jahre eine Umwandlung der kindlichen Unbekümmertheit und Begeisterungsfähigkeit in den nüchternen, vorsichtig-zurückhaltenden Ernst herbeiführten, der nach den Ergebnissen Cattells und seiner Mitarbeiter für kreative Erwachsene kennzeichnend ist.***	Die Hypothese wird durch die Untersuchungsergebnisse von Werner und Bachtold (1969) für 13- bis 15jährige Mädchen bestätigt, alle anderen Resultate der 3 Untersuchungen mit dem CPQ sind für F nicht signifikant **. Die Charakterisierung der Kinder mit hohen F-Werten entspricht derjenigen der kreativen und intelligenten Kinder in der Untersuchung von Wallach und Kogan (1965)**.
G	Unzuverlässig, impulsiv, geringe Überich-Stärke versus gewissenhaft, ausdauernd, hohe Überich-Stärke. Dieser Faktor ist insofern schulleistungsrelevant, als zu den verinnerlichten gesellschaftlichen Normen auch eine hohe Leistungsmotivation gehört.	-	-	-	-

*) Vgl. Tabelle III in Anhang B 2

**) S. z.B. Wallach und Kogan S. 86-93, vor allem S. 89. Die Kreativitätseffekte waren zwar nur für die Mädchen signifikant, aber die Charakterisierung der kreativen und intelligenten Jungen stimmt mit derjenigen der Mädchen recht gut überein.

***) Nach den Informationen der Tabelle III könnte angenommen werden, daß bei Jungen kein positiver Zusammenhang zwischen Kreativität und den F-Scores besteht oder daß die beschriebene Veränderung bei Jungen früher in Gang gesetzt wird. Diese Annahme kann jedoch nicht durch einen Hinweis auf bestimmte Merkmale der männlichen Geschlechtsrolle gestützt werden. Es wäre im Gegenteil erstaunlich, wenn die soziale Umwelt bei kreativen Jungen zwar Dominanz und Aggressivität, nicht jedoch eine lebhafte, fröhliche Unbekümmertheit dulden würde.

Tabelle 2o (Fortsetzung)

Faktor		Hypothese	Begründung der Hypothese		
Bezeichnung	Beschreibung		Kreativitätstheorie(n)	Spezifikationsgleichung	Untersuchung von
H	Ängstlich, schüchtern, gehemmt versus beherzt, spontan, ungehemmt. Diese Komponente der Extraversions-Introversions-Dimension umfaßt die emotionalen Reaktionen auf die soziale Umwelt. Kinder mit hohen Scores für H interagieren frei und ungehemmt mit anderen, während Kinder mit niedrigen H-Werten solche Interaktionen oft als bedrohlich empfinden und mit Rückzugsverhalten reagieren.	Zwischen den Ergebnissen der Versuchsform VI, K und den H-Scores besteht ein positiver Zusammenhang.	Wie A und F ist auch H eine Komponente des Faktors zweiter Ordnung "Introversion", der bei kreativen Persönlichkeiten "eine gewisse paradoxe Struktur"* besitzt, weil H, der in der Gesamtpopulation mit A und F positiv korreliert, bei Kreativen mit diesen beiden Faktoren negativ korreliert. Diese bei kreativen Erwachsenen gefundene negative Beziehung ist nach der hier vorgeschlagenen Begründung der Hypothese darauf zurückzuführen, daß die hohen F-Werte der kreativen Kinder infolge negativer Umweltreaktionen auf bestimmte Persönlichkeitsmerkmale (eventuell auf den stärker hereditär determinierten Faktor H und auf Faktor E) im Lauf der Entwicklung zurückgingen, während bei kreativen Kindern die positive Korrelation noch besteht.	+ .1 H	Porter 1964. Werner 1966, Werner und Bachtold 1969 sowie die Ergebnisse von Wallach und Kogan (1965) für ihre Gruppe kreativer und intelligenter Mädchen.
I	Robust, realistisch, unabhängig versus sensibel, phantasievoll, ängstlich gegenüber körperlicher Bedrohung, abhängig. Die Erziehung von Kindern mit hohen I-Werten ist durch ein hohes Maß an "overprotection" gekennzeichnet.	Zwischen den Ergebnissen der Versuchsform VI, K und den Werten für den Faktor I besteht eine negative Beziehung.	Die Folge einer übermäßig behütenden Erziehung ist in erster Linie Abhängigkeit. Daß Kinder mit niedrigen I-Werten eher zu kreativen Leistungen prädestiniert sind, kann aus der großen Bedeutung der Unabhängigkeit als Voraussetzung für schöpferisches Verhalten gefolgert werden.**	- .1 I	Porter (1964) erhielt für 31 Jungen ein Resultat, das die Hypothese stützt, während Werner und Bachtold für eine Stichprobe von 113 Jungen ein der Hypothese widersprechendes Ergebnis mitteilen.
J	Begeistungsfähig, aktiv, unkritisch, Vorliebe für Gruppenaktivitäten versus vorsichtig, (selbst-) kritisch, wählerisch, individualistisch.	-		-	
N	Offen, ungekünstelt, gefühlvoll versus schlau, berechnend, opportunistisch. Bei Kindern ist die Bedeutung dieses Faktors weniger klar definiert; Kinder mit hohen N-Werten erscheinen infolge eines "verständigen", gewitzten Sozialverhaltens besser in der Lage, ihre eigenen Interessen wahrzunehmen als Kinder mit niedrigen Werten für diesen Faktor.				

*) Cattell und Butcher 1968, S. 278

Tabelle 2o (Fortsetzung)

Faktor		Hypothese	Begründung der Hypothese		
Bezeichnung	Beschreibung		Kreativitätstheorie(n)	Spezifikationsgleichung	Untersuchung von
Ø	Selbstsicher, gelassen, heiter, ohne Schuldgefühle versus unsicher, ängstlich, besorgt, zu Schuldgefühlen und Skrupeln neigend. Der Faktor O differenziert bei Erwachsenen am besten zwischen Neurotikern und der Gesamtpopulation. Die Persönlichkeitsstörungen der Individuen mit hohen Punktwerten können verschiedene Formen annehmen, z. B. Angst, Depressionen, Schuldgefühle u.s.	Zwischen den Ergebnissen der Versuchsform VI, K und den Scores für O besteht eine negative Beziehung.	Die Ich-Stärke und emotionale Stabilität der kreativen Kinder ist mit relativ niedrigen Werten für den Faktor O verbunden.	- .1 Ø	Porter 1964, Werner 1966, Werner und Bachtold 1969.
Q_3	Willensschwach, nachlässig, unordentlich, unbekümmert um soziale Normen versus willensstark, kontrolliert, diszipliniert, zwanghaft. Kinder mit niedrigen Scores für diesen Faktor können durch Unbedachtsamkeit und Nachlässigkeit häufig gegen schulische Regeln verstoßen.	-	-	-	-
Q_4	Entspannt, ruhig, apathisch, geringe Triebspannung versus angespannt, erregbar, ärgerlich, hohe Triebspannung. Kinder mit hohen Q_4-Scores haben eine geringe Frustrationstoleranz und neigen zu Wutausbrüchen.	-	-	-	-

Da der CPQ nur übersetzt, aber nicht für den Gebrauch mit deutschen Kindern adaptiert ist, wäre für den Fall überwiegend negativer Ergebnisse dieser Hypothesenüberprüfung keine Entscheidung darüber möglich, ob der CPQ in dieser Form für deutsche Kinder nicht valide ist oder ob die Versuchsform IV, K etwas anderes als Kreativität mißt. Aus diesem Grund werden die Ergebnisse des Kreativitätstests von Wallach und Kogan, dessen erste Validitätsuntersuchungen mit deutschen Stichproben überwiegend positive Ergebnisse erbrachten,[1] in die Überprüfung der Hypothesen einbezogen. Wenn die vermuteten Zusammenhänge mit den CPQ-Faktoren weder bei der Versuchsform VI, K noch bei dem Verfahren von Wallach und Kogan auftreten, könnte dies nicht als Hinweis auf die geringe Validität der beiden Kreativitätstests interpretiert werden, da die alternative Erklärung dieses Resultats durch die fehlende interkulturelle Generalisierbarkeit der CPQ-Ergebnisse zumindest ebenso wahrscheinlich wäre. In diesem Fall wäre also der Versuch einer Validitätsüberprüfung der Versuchsform VI, K mit Hilfe des CPQ gescheitert.

Die Ergebnisse der beiden Kreativitätstests sollten durch die Skalenwerte des CPQ mit Hilfe schrittweiser Regressionsanalysen für Jungen und Mädchen getrennt vorhergesagt werden. Wenn die Faktoren B, C und E bei Jungen die Ergebnisse der Versuchsform VI, K am besten vorhersagen, sollte dies als Validitätshinweis für den Kreativitätstest gewertet werden. Auch die Bestätigung einer oder mehrerer Hypothesen zu den Faktoren F, H, I und O soll als Beleg für die Konstruktvalidität der Versuchsform VI, K interpretiert werden. Falls weder B, C noch E zu den 3 besten Prädiktoren dieser Testergebnisse gehören sollten, müßte das als negatives Resultat dieses Teils der Validitätsuntersuchung betrachtet werden.

Nachdem die grundlegenden Untersuchungen Cattells und seiner Mitarbeiter an männlichen Kreativen durchgeführt wurden[2] und nachdem Werner (1966) und Werner und Bachtold (1969) für kreative Mädchen wesentlich geringere Übereinstimmungen mit den Persönlichkeitsprofilen kreativer Erwachsener fanden, werden für Mädchen weniger enge Zusammenhänge zwischen den CPQ-Faktoren und den Ergebnissen der Versuchsform VI, K erwartet.

---

1) S. dazu S. 1o5-11o

2) Z.B. Cattell und Drevdahl (1955) und Drevdahl und Cattell (1958)

Die vorangegangenen Überlegungen zu der Interpretation der Ergebnisse dieses Teils der Validitätsprüfung sind insofern unbefriedigend, als sie nicht für jedes mögliche Resultat festlegen, ob es als Validitätsbeleg gewertet werden soll oder nicht. Zugleich erscheint es jedoch aufgrund der oben dargestellten Probleme und Schwierigkeiten dieser Untersuchung zur Konstruktvalidität der Versuchsform VI, K nicht angezeigt, die Hypothesen noch stärker zu spezifizieren.
Als Ausweg bietet sich ein Vergleich der Vorhersagbarkeit der Kreativitätstestwerte und der P-S-B-Ergebnisse durch die CPQ-Faktoren an. Wegen der stärkeren Beeinflußbarkeit des divergenten Denkens durch nichtintellektuelle Persönlichkeitsfaktoren wird erwartet, daß die Gesamtwerte der Versuchsform VI,K und des Kreativitätstests von Wallach und Kogan durch die CPQ-Faktoren besser vorhergesagt werden als die P-S-B-Ergebnisse. Auch diese Hypothese wird nur für die Jungen aufgestellt.

3.33 Darstellung und Diskussion der Ergebnisse

Zunächst werden die Untersuchungsergebnisse der Gesamtstichprobe, dann diejenigen der Teilstichprobe besprochen.

3.331 Ergebnisse der Gesamtstichprobe

Obwohl eine Reihe von korrelativen Zusammenhängen aufgrund der Ergebnisse der Gesamtstichprobe untersucht wurde, erscheint es nicht angezeigt, die statistischen Kennwerte dieser 4 Schultypen umfassenden Stichprobe mitzuteilen. Dagegen sind die schultypspezifischen Mittelwerte, Streuungen und die Ergebnisse der Normalitätsprüfung u.U. von Interesse, sie sind im Anhang dargestellt (s. Anhang B 3).

3.3311 Diskriminante Validität des Verfahrens

Erstes Ziel der Untersuchung in der Gesamtstichprobe war die Überprüfung der diskriminanten Validität der Versuchsform VI, K.[1)]
Da die Korrelation zwischen dem Gesamtmaß für Produktivität und dem P-S-B-Gesamtergebnis .43 und diejenige zwischen dem Produktivitäts- und dem BBT-Gesamtwert .44 beträgt, erscheint der Nachweis, daß die Aufgaben zur Erfassung des divergenten Denkens untereinander nicht nur aufgrund ihrer Beziehungen zur Testintelligenz zusammenhängen, besonders wichtig.

1) Vgl. dazu S. 198

Die über z-Transformation berechnete mittlere Korrelation der Anzahl- und Originalitätswerte der 11 Aufgaben beträgt .38;[1] wenn die Koeffizienten für die Beziehung zwischen Produktivität und Originalität bei derselben Aufgabe nicht in die Durchschnittsberechnung einbezogen werden, beträgt die mittlere Korrelation .345.
Tabelle 21 zeigt die gemittelten Korrelationen der Aufgaben der Versuchsform VI, K untereinander und mit den Subtests des Prüfsystems für Schul- und Bildungsberatung von Horn (P-S-B) und mit den Subtests des im Zentrum für empirische pädagogische Forschung entwickelten Bildungs-Beratungs-Tests (BBT).

Tabelle 21: Mittlere Interkorrelationen der Items der Versuchsform VI, K untereinander und mit den Subtests des P-S-B und des BBT (Koeffizienten der Gesamtstichprobe)

	VI, K	P-S-B	BBT
VI, K	$\bar{r}$ = .345 N = 1671	$\bar{r}$ = .16 N = 1205	$\bar{r}$ = .21 1464 ≤ N ≤ 1487

Die Items der Versuchsform VI, K korrelieren untereinander wesentlich höher als mit den Subtests der beiden Verfahren zur Messung der Intelligenz. Die Signifikanz der Unterschiede der 3 mittleren Korrelationskoeffizienten kann allerdings nicht überprüft werden.[2]

Sämtliche Koeffizienten sind höher als die entsprechenden Werte der Versuchsform IV, 1.[2] Um zu überprüfen, ob dies allein auf die Erhöhung der Testwertstreuung durch die Verarbeitung der Daten verschiedener Schultypen zurückzuführen ist, wurden auch die mittleren Korrelationen der Schüler der Gesamtschulen dieser Stichproben berechnet (s. Tabelle 22).

---

1) Trotz der signifikanten Abweichungen von der Normalverteilung wurden Produkt-Moment-Korrelationen berechnet (s. dazu Anmerkung 1, S. 155). Der Einfluß von Abweichungen von der Normalverteilung auf die Berechnung von Produkt-Moment-Korrelationen scheint nicht sehr stark zu sein. Guilford z.B. hält die Normalverteilung der Meßwerte nicht für unbedingt erforderlich (Guilford 1965, zitiert nach Heller und Rosemann 1974).

2) Vgl. dazu S. 156

Tabelle 22: Mittlere Interkorrelationen der 21 Scores der Versuchsform VI, K untereinander und mit den Subtests des P-S-B und des BBT (Koeffizienten der Gesamtschul-Stichprobe)

	VI, K	P-S-B	BBT
VI, K	$\bar{r}$ = .32*	$\bar{r}$ = .15	$\bar{r}$ = .16
	N = 873	N = 5oo	756 ≤ N ≤ 763

*) Wenn die 1o Koeffizienten für die Korrelation zwischen Anzahl und Originalität derselben Aufgabe in die Durchschnittsberechnung einbezogen werden, ergibt sich eine Korrelation von .35.

Da diese Zahlen in etwa denjenigen der Tabelle 7 entsprechen,[1] kann angenommen werden, daß die z.T. höheren Koeffizienten der Gesamtstichprobe vor allem auf ihre stärkere Testwertstreuung zurückzuführen sind. Die geringfügigen Erhöhungen der durchschnittlichen Interkorrelationen der Kreativitäts-Items von .3o auf .32 sowie des $\bar{r}$ für die Kreativitäts-BBT-Beziehung von .14 auf .16 müssen als Zufallsschwankungen betrachtet werden. Die Erwartung, daß die Items der Versuchsform VI, K eine höhere interne Konsistenz besitzen als diejenigen der Versuchsform VI, 1, hat sich nicht erfüllt, die interne Konsistenz hat sich nur unwesentlich erhöht.

Sie erscheint allerdings durchaus befriedigend, wenn man berücksichtigt, daß es sich um die durchschnittliche Korrelation einzelner Items handelt. Die 5 Produktivitäts- und die 5 Originalitätswerte des Kreativitätstests von Wallach und Kogan interkorrelieren zwar höher - in der Untersuchung der Testautoren beträgt die durchschnittliche Korrelation .43, in derjenigen von Krause (1972) .42 (Berechnung über z-Transformation) - aber wenn man die höhere Reliabilität der aus 4 - 1o Items bestehenden 5 Subtests des Verfahrens von Wallach und Kogan in Rechnung stellt, schneidet das $\bar{r}$ der Versuchsform VI, K von .32 sehr gut ab.
Die Kreativitäts-Intelligenz-Korrelationen sind in allen Untersuchungen deutlich niedriger als die Interkorrelationen der Items der Versuchsform VI, K.

Außer diesem Vergleich der Korrelationsmittelwerte wurden zur Überprüfung der diskriminanten Validität der Versuchsform VI, K auch Faktorenanalysen gerechnet. Da neben den 11 Anzahl- und 1o Originalitätswerten der Versuchsform VI, K, den P-S-B- und BBT-Werten und den

1) Vgl. S. 156

4 Ergebnissen des Angstfragebogens für Schüler[1] auch Schulleistungsvariablen in Form von Zensuren[2] in die Berechnungen eingingen, mußten die Analysen für die 4 Schultypen getrennt durchgeführt werden. Außer den genannten Variablen wurden noch das Alter der Schüler, Beruf und Schulbildung des Vaters und die Schulbildung der Mutter berücksichtigt.

Da bei insgesamt 55 Untersuchungsvariablen der Anteil an Schülern mit **missing data ziemlich hoch war und das Programm PAFA Probanden mit** missing data nicht berücksichtigt, beruhen nur die Faktorenanalysen der Realschul- und Gesamtschuldaten auf einer ausreichenden Zahl von Versuchspersonen.

Die Eigenwertgrenze wurde auf 1.oo festgesetzt.

Von den Faktorenlösungen, die dem Fürntratt-Kriterium genügen[3], werden für beide Schultypen die 5-Faktoren-Lösungen, die psychologisch interpretierbare Faktoren beinhalten, in den folgenden Tabellen 23 und 24 wiedergegeben.

Bei den Gesamtschuldaten werden durch die 5 Faktoren 5o,1 Prozent der Gesamtvarianz aufgeklärt, bei den Realschuldaten 42,6 Prozent.

---

1) Der AFS untersucht die Faktoren Prüfungsangst, allgemeine (manifeste) Angst, Schulunlust und die Tendenz, sich angepaßt und sozial erwünscht darzustellen.

2) Die Zensuren in Deutsch, Mathematik, erster Fremdsprache und Kunsterziehung (s. dazu S. 197)

3) Jeder Faktor einer Faktorenlösung soll durch mindestens 3 Variablen repräsentiert sein, für die der Quotient aus Ladungsquadrat und quadrierter Kommunalität größer als .5o ist (vgl. Fürntratt 1969).

Tabelle 23: Rotierte Faktormatrix[1] ausgewählter Variablen der Untersuchung in Gesamtschulen (N = 163).

Faktoren :	1.	2.	3.	4.	5.	h^2
Alter	.29	.08	.09	.05	.02	.17
P-S-B Untertest 1 u.2	.70	-.20	.04	-.01	.18	.56
" " 3	.62	.08	.07	-.04	-.11	.41
" " 4	.72	.02	.04	-.08	.00	.53
" " 5	.27	-.18	.22	-.25	.23	.27
" " 6	.68	-.04	.10	-.06	.02	.47
" " 7	.43	-.25	.10	-.13	-.10	.28
" " 8	.51	-.17	-.04	-.08	-.01	.30
" " 9	.37	-.24	.11	-.10	.16	.24
" " 10	-.28	.40	.08	-.08	-.02	.25
BBT Zahlenschrift	.57	.08	.19	-.24	-.21	.47
" Wortbedeutungen	.52	.15	.01	.32	.20	.43
" Satzteile	.55	.19	-.03	.22	.15	.41
" Zahlenreihen	.46	-.08	-.04	.43	-.09	.42
" Denkaufgaben	.60	.03	-.02	.40	.05	.52
Zensur Deutsch 1	.54	.15	-.03	.48	.35	.66
" " 2	.54	.12	.03	.54	.36	.73
" " 3	.32	.01	-.01	.72	-.16	.68
" Mathematik 1	.60	-.04	-.01	.58	-.01	.70
" " 2	.50	-.01	-.09	.63	-.05	.66
" " 3	.33	-.10	-.16	.69	.00	.62
" Fremdsprache 1	.62	.07	-.03	.36	-.05	.53
" " 2	.65	.03	.02	.52	-.02	.70
" " 3	.43	.04	-.09	.57	-.03	.53
" Kunsterzieh. 1	.20	.18	.02	.48	-.17	.33
" " 2	.13	.18	-.05	.65	-.12	.49
" " 3	.18	.11	.06	.57	-.09	.38
AFS Prüfungsangst	.23	.04	.12	.60	.13	.44
" Manif. Angst	.07	.02	.14	.38	.19	.21
" Schulunlust	.07	.21	.09	.34	.10	.18
" Soz. Erwünschth.	.18	-.22	-.17	.02	-.01	.11
Beruf des Vaters	-.03	.21	.17	.59	-.12	.44
Bildung des Vaters	.03	.27	.04	.63	-.04	.48
" der Mutter	-.02	.25	.10	.67	-.17	.54
VI, K 1 Anzahl	.01	-.06	.32	-.21	.25	.21
" 2 "	-.06	-.11	.18	.16	.74	.63
" 2 Originalität	-.09	-.08	.10	.17	.75	.61
" 3 Anzahl	.08	-.27	.19	-.05	.64	.53
" 3 Originalität	.03	-.25	.20	-.08	.64	.52
" 4 Anzahl	.01	-.16	.85	-.13	-.02	.77
" 4 Originalität	-.07	-.17	.79	-.05	.01	.66
" 5 Anzahl	-.08	-.13	.85	-.06	.17	.78
" 5 Originalität	-.10	-.10	.67	.15	.25	.56
" 6 Anzahl	-.03	-.26	.81	.06	.18	.76
" 6 Originalität	-.01	-.16	.74	.06	.12	.60
" 7 Anzahl	-.04	-.70	.02	-.11	.24	.56
" 7 Originalität	-.20	-.63	.05	-.15	.21	.50
" 8 Anzahl	-.12	-.56	.09	-.17	.30	.45
" 8 Originalität	-.17	-.61	.10	-.04	.30	.50
" 9 Anzahl	-.01	-.78	.17	.09	.09	.66
" 9 Originalität	.01	-.72	.06	.05	.06	.52
" 10 Anzahl	.00	-.69	.44	.13	.03	.68
" 10 Originalität	.03	-.64	.36	.03	.05	.54
" 11 Anzahl	.00	-.81	.11	.09	.01	.68
" 11 Originalität	.04	-.79	.15	.07	-.02	.65
% der durch die 5 Faktoren aufgeklärten Varianz	25.25	22.02	16.97	24.38	11.38	

1) Programm PAFA des DRZ; s. dazu Anmerkung 1, S. 156

Die Ergebnisse der Faktorenanalyse der Gesamtschuldaten bieten wie in der Voruntersuchung ein klares Bild. Auf dem ersten Faktor laden alle BBT- und 8 der 1o P-S-B-Untertests, die Zensuren in Deutsch, Mathematik und erster Fremdsprache sowie das Alter mit Ladungszahlen über .3o; er könnte als "schulleistungsrelevante intellektuelle Fähigkeiten" interpretiert werden. Die Ladungszahlen der Noten werden vom ersten Halbjahreszeugnis der 5. bis 1. Halbjahreszeugnis der 8. Klasse in allen 3 Hauptfächern geringer, was eventuell im Sinn eines abnehmenden Einflusses von Intelligenzfaktoren auf die Schulleistungen oder Schulleistungsbeurteilungen zu erklären ist.
Der 4. Faktor, der für die Schulleistungen ebenfalls von großer Bedeutung ist, wird dagegen im Verlauf der ersten 3 Halbjahre immer wichtiger für die Zensuren der 3 Hauptfächer. Da auf ihm die 3 Variablen des sozialen Status und - mit entgegengesetztem Vorzeichen- die Ergebnisse des Angstfragebogens, vor allem die Prüfungsangst-Skala, hoch laden, wird er als ein vom Bildungsniveau des Elternhauses abhängiger Motivationsfaktor interpretiert. Ob die 3 Untertests des BBT infolge einer gewissen Beeinflußbarkeit der Ergebnisse durch Prüfungsangst auf diesem Faktor laden oder ob die 3 Subtests einen schichtspezifischen Bias haben, kann hier nicht entschieden werden.
Sämtliche 21 Variablen der Versuchsform VI, K laden nicht auf diesen beiden, sondern auf dem 2., 3. und 5. Faktor. Diese Aufspaltung auf 3 Faktoren deutet auf die faktorielle Komplexität des Verfahrens hin, die sich bereits bei der 3-Faktoren-Lösung abzeichnet. (Während Intelligenz-, Schulleistungs-, Angst- und Statusvariablen auf dem ersten Faktor laden,[1)] verteilen sich die substantiellen Ladungen der 21 TDK-Resultate auf den 2. und 3. Faktor.)
Auf dem 2. Faktor laden verbal zu beantwortende Aufgaben und zwar solche mit verbalen (Aufgabe 7 und 8, "lustige Vergleiche") und solche mit visuellen Aufgabenstimuli (9 - 11, "Bilderraten"). Außer diesen 5 Aufgaben der Versuchsform VI, K lädt auf dem 2. Faktor nur der Untertest 1o des P-S-B mit einer vergleichsweise niedrigen Ladung von .4o. (An sich war erwartet worden, daß Untertest 5, der Guilfords Faktor "divergente Produktion symbolischer Einheiten" DSU mißt, die höchste Ladung auf dem 2. Faktor besitzt; die Ladungen dieses Untertests verteilen sich jedoch relativ gleichmäßig auf alle 5 Faktoren.)
Der 3. Faktor wird ausschließlich durch VI, K-Variablen repräsentiert und zwar nur durch rein oder überwiegend zeichnerisch zu lösende Items

---

1) Nur der Subtest 1o des P-S-B ist auszunehmen, er lädt am höchsten auf dem 2. Faktor.

(1, "Spielklötze", und 4 - 6, "Zeichnungen"); er könnte daher als Faktor des divergenten Denkens im figuralen Inhaltsbereich bezeichnet werden.
Auf dem 5. Faktor laden die beiden Aufgaben, in denen die Kinder Konsequenzen irrealer Situationen aufschreiben sollen (Aufgabe 2 und 3) sowie - mit .3o - die von Wallach und Kogan übernommene Aufgabe 8 ("Lustige Vergleiche" zwischen Katze und Maus), die ihre höchsten Ladungen[1)] auf dem 2. Faktor hat. Dieser Faktor des divergenten Denkens im semantischen Inhaltsbereich ist für die beiden Deutschzensuren der 5. Klasse von Bedeutung.

1) Für Anzahl und Originalität

Tabelle 24: Rotierte Faktormatrix ausgewählter Variablen der Untersuchung in Realschulen (N = 243)

	1. Faktor	2. Faktor	3. Faktor	4.Faktor	5.Faktor	Kommunalität
Alter	.14	.17	.37	-.o3	-.o4	.19
P-S-B Untertest 1 u. 2	.17	.21	.64	.o1	.1o	.49
" " 3	.o5	-.o2	.61	.o1	.o3	.38
" " 4	.22	.o5	.68	-.13	-.o4	.53
" " 5	.o5	.36	.37	-.26	.1o	.35
" " 6	.17	.o8	.58	-.o6	-.o3	.38
" " 7	.o9	-.o1	.51	-.11	-.o6	.29
" " 8	.o9	-.17	.41	.o3	.o3	.2o
" " 9	.14	.11	.45	-.15	.oo	.26
" " 1o	-.o1	.15	.47	-.15	-.12	.28
BBT Zahlenschrift	.32	o8	-.29	-.o2	-.o6	.2o
" Wortbedeutungen	.39	.23	-.26	.11	.o4	.29
" Satzteile	.4o	.17	.36	.oo	-.15	.34
" Zahlenreihen	.23	.16	.4o	-.1o	-.o8	.26
" Denkaufgaben	.35	.23	.45	.o2	.o7	.38
Zensur Deutsch 1	.63	-.2o	.39	.o4	-.14	.61
" " 2	.68	-.18	.34	.11	-.14	.64
" " 3	.78	.o2	.19	-.o2	-.o8	.65
" Mathematik 1	.68	.1o	.32	.o2	.o3	.58
" " 2	.77	.o8	.23	-.o1	-.o2	.65
" " 3	.68	.o8	.15	.o2	.o1	.49
" Fremdsprache 1	.73	-.o6	.o4	-.17	.o2	.57
" " 2	.77	.1o	.o2	-.21	-.o2	.65
" " 3	.8o	-.o2	.12	-.15	-.o8	.69
" Kunsterzieh. 1	.48	.o5	-.o1	.25	-.o6	.3o
" " 2	.43	-.o4	-.11	.29	.o9	.29
" " 3	.52	.o2	-.o4	.28	-.o1	.35
AFS Prüfungsangst	-.24	.oo	.2o	-.o6	.34	.21
" Manif. Angst	-.o6	-.o6	.o2	-.15	.45	.23
" Schulunlust	-.2o	-.o1	-.o2	.o6	.16	.o7
" Soz. Erwünschth.	-.o3	-.o7	.13	-.o6	-.12	.o4
Beruf des Vaters	.o7	.14	-.22	.2o	.o5	.11
Bildung des Vaters	.o8	.22	.39	.15	.o6	.23
Bildung des Mutter	.o6	.23	-.19	.o8	.o5	.1o
VI, K 1 Anzahl	-.o2	.31	-.15	.o7	.o9	.13
" 2 "	.oo	.18	-.o9	-.o2	.79	.66
" 2 Originalität	.o4	.14	-.o9	-.o2	.75	.59
" 3 Anzahl	.19	.29	-.o2	-.18	.58	.49
" 3 Originalität	.23	.21	.13	-.16	.61	.51
" 4 Anzahl	-.o6	.62	-.o3	-.15	.o7	.42
" 4 Originalität	-.o9	.55	-.o3	-.1o	.1o	.33
" 5 Anzahl	.o1	.67	-.1o	-.35	.13	.59
" 5 Originalität	.o5	.49	.oo	-.2o	.14	.3o
" 6 Anzahl	.o1	.72	.12	-.32	-.o2	.63
" 6 Originalität	.o2	.68	.2o	-.28	-.o8	.59
" 7 Anzahl	.o8	.63	-.1o	-.21	.o2	.46
" 7 Originalität	.o8	.63	-.o9	-.o6	-.o5	.41
" 8 Anzahl	-.o2	.65	-.18	-.o5	.o6	.46
" 8 Originalität	.o3	.67	-.2o	.o2	.17	.51
" 9 Anzahl	.o8	.3o	-.o8	-.81	.o3	.76
" 9 Originalität	.o4	.28	-.o2	-.78	.o9	.69
" 1o Anzahl	.oo	.19	-.o4	-.73	.21	.61
" 1o Originalität	.oo	.11	-.o7	-.74	.19	.6o
" 11 Anzahl	.o2	.25	-.15	-.77	.o1	.68
" 11 Originalität	.o1	.19	-.13	-.78	.o3	.66
% der durch die Faktoren aufgeklärten Varianz	27.58	22.52	18.92	19.84	11.14	

Auch für die Realschuldaten ergibt sich trotz der z.T. abweichenden Ergebnisse die gleiche klare Trennung zwischen Intelligenztest-, Schulleistungs- und Statusvariablen einerseits und den VI, K-Ergebnissen andererseits. Letztere laden auf dem 2., 4. und 5. Faktor, auf denen außer dem 5. Untertest des P-S-B kein Intelligenz-Subtest und keine Zensursubstantielle Ladungen besitzt. Auf dem 5. Faktor laden nur die 2. und 3. Aufgabe ("Verrückte Geschichten") und außerdem - mit gleichem Vorzeichen, aber niedrigeren Ladungszahlen - die Skalen "Prüfungsangst" und "manifeste Angst" des AFS, was eventuell so zu interpretieren ist, daß ängstlichere Realschüler zu diesen beiden Aufgaben mehr Antworten aufschreiben als wenig ängstliche.[1)]

Zusammenfassend kann man die faktorenanalytischen Ergebnisse für beide Schultypen folgendermaßen beschreiben: Die Produktivitäts- und Originalitätswerte der Versuchsform **VI**, K laden auf 3 Faktoren am höchsten, auf denen die anderen Variablen - abgesehen von jeweils 3 der insgesamt 34 Variablen - Ladungen unter .3o besitzen. Damit wird die faktorielle Abgrenzung der VI, K-Werte von den Intelligenztestvariablen durch die Ergebnisse der Gesamt- und der Realschulstichprobe bestätigt.
Dieses Resultat erhält durch die faktorielle Komplexität der Versuchsform VI, K sowie durch die Abweichungen der Faktorenstruktur der beiden Stichproben besonderes Gewicht. Es kann als Beleg für die diskriminante Validität der Versuchsform VI, K gewertet werden.

Somit wurden die Hypothesen, die zu der Untersuchung der diskriminanten Validität der Versuchsform VI, K aufgestellt worden waren[2)], im wesentlichen bestätigt. Lediglich die Erhöhung der internen Konsistenz des Verfahrens zeigte sich nicht in dem erwarteten Ausmaß.

3.3312 Geschlechtsabhängigkeit der Testergebnisse

Ein weiteres Ziel der Untersuchung in der Gesamtstichprobe war die Überprüfung der Geschlechtsabhängigkeit der Versuchsform VI, K. Tabelle 25 zeigt die Ergebnisse des t-Tests für unabhängige Stichproben [3)] für die Jungen und Mädchen aus den Gesamtschulen.

---

1 Die z.T. sehr interessanten Abweichungen der Faktorenstruktur der Gesamt- und der Realschuldaten werden nicht besprochen, da sie für die Untersuchungsfrage irrelevant sind.

2) S. dazu S.198

3) Programm PAMV des Deutschen Rechenzentrums, Autor: Dr. F. Gebhardt; DRZ Darmstadt

Tabelle 25: Arithmetische Mittel und t-Werte für die Mittelwertsdifferenzen zwischen den Testergebnissen von 41o Jungen und 432 Mädchen aus Gesamtschulen

VI, K Aufgabe	$\bar{X}$ Jungen	$\bar{X}$ Mädchen	t
1	6,32	6,11	o,8o
2 Anzahl	4,38	4,37	o,o5
3 "	3,87	4,3o	3,14 ss*
4 "	5,o7	5,47	2,7o ss
5 "	4,23	4,68	3,o1 ss
6 "	3,88	4,25	2,69 ss
7 "	4,23	4,46	1,75
8 "	5,3o	5,76	2,68 ss
9 "	5,93	7,21	6,13 ss
1o "	4,47	5,23	4,63 ss
11 "	5,47	6,4o	4,46 ss
2 Originalität	6,13	6,o5	o,34
3 "	4,41	4,8o	1,84
4 "	6,58	6,87	1,24
5 "	4,95	4,79	o,67
6 "	5,13	5,81	2,97 ss
7 "	3,55	3,56	o,o8
8 "	5,24	5,42	o,74
9 "	6,o7	7,26	3,72 ss
1o "	3,26	3,78	2,14 s
11 "	5,53	6,8o	3,81 ss
Anzahl-Gesamtwert	53,15	58,24	4,3o ss
Originalitäts-Gesamtwert	5o,85	55,14	2,88 ss
Gesamt-Gesamtwert	1o4,oo	113,38	3,58 ss

*) Signifikante t-Werte ($p \leq .o5$) sind durch s, sehr signifikante ($p \leq .o1$) durch ss gekennzeichnet (zweiseitiger Test).

Aus der Tabelle 25 geht klar hervor, daß die überwiegende Zahl der Items die Mädchen begünstigt. Die Mädchen erreichen wie in der Untersuchung von Bhavnani und Hutt (1972)[1] häufiger bei den Produk-

1) S. dazu S. 136-138

tivitäts- als bei den Originalitätswerten bessere Resultate als die Jungen.
Da die Mädchen jedoch auch in 4 von 1o Aufgaben höhere Originalitätsscores erhalten als die Jungen, sollten für das Verfahren geschlechtsspezifische Normen entwickelt werden.
Der Gefahr, daß das Verfahren für Mädchen weniger valide ist als für Jungen, wird durch eine geschlechtsspezifische Normierung allerdings nicht begegnet, da nicht anzunehmen ist, daß die miterfaßte Variable sich wie eine additive Konstante auf die Testwerte auswirkt. Wenn es sich dabei wirklich um schulische Anstrengungsbereitschaft handeln würde, wäre z.B. zu erwarten,daß diese sich bei intelligenteren Mädchen und bei guten Schülerinnen im allgemeinen stärker auswirkt als bei weniger intelligenten Mädchen und bei schlechten Schülerinnen, da die letzteren für eigene Anstrengungen seltener belohnt werden als erstere.
Damit dieser eventuellen Beeinträchtigung der Validität entgegengewirkt wird, müßte sich das Verfahren u.a. durch Instruktion und Untersuchungssituation von schulischen Leistungsprüfungen stark unterscheiden, so daß die Kinder keinen Anlaß sehen, sich aus schulleistungsorientierten Motivationen bei der Testbearbeitung besonders anzustrengen. Im Übrigen sollte das Problem der differentiellen Validität des Verfahrens im Auge behalten werden.

3.3313 Schichtabhängigkeit der Testergebnisse

Ein erster Hinweis zu der Frage der Schichtabhängigkeit der Versuchsform **VI**, K kann den Ergebnissen der beiden Faktorenanalysen entnommen werden: Danach sind weder der Väterberuf noch die Schulbildung der Eltern für die Variablen des divergenten Denkens von Bedeutung.[1)]
Die Korrelationskoeffizienten für den Zusammenhang zwischen den Testergebnissen und den 4 Maßen des Sozialstatus[2)] werden in Tabelle 26 dargestellt.

---

1) S. Tabelle 23, S.217,und Tabelle 24, S. 22o
2) S. dazu S. 197

Tabelle 26: Produkt-Moment-Korrelationen[1] zwischen den Produktivitäts- und Originalitätswerten der VI, K-Aufgaben und 4 Variablen des Sozialstatus für die Gesamtstichprobe

Versuchsform VI, K	Vaterberuf N = 464	Schulbildung des Vaters N = 451	Schulbildung der Mutter N = 459	Sozialstatus nach Bauer N = 464
Aufgabe 1 Anzahl	.15 ss*	.19 ss	.17 ss	.19 ss
" 2 "	.00	.00	-.03	-.01
" 3 "	-.01	.00	-.05	-.02
" 4 "	.14 ss	.10	.13 ss	.14 ss
" 5 "	.06	.06	.10 s	.08
" 6 "	.02	.04	-.02	.02
" 7 "	.07	.01	-.03	.03
" 8 "	.04	.01	.01	.03
" 9 "	-.04	-.05	-.07	-.06
" 10 "	.00	-.02	-.04	-.02
" 11 "	-.06	-.05	-.06	-.06
" 2 Originalität	-.01	.00	-.04	-.02
" 3 "	.02	.01	-.03	.00
" 4 "	.10	.08	.09	.11 s
" 5 "	-.01	-.06	-.03	-.03
" 6 "	.00	.05	-.01	.02
" 7 "	.12 s	.04	-.01	.07
" 8 "	-.02	-.04	-.03	-.03
" 9 "	-.07	-.05	-.07	-.07
" 10 "	.06	.03	.00	03
" 11 "	-.07	-.05	-.06	-.07
Produktivitäts-Gesamtwert	.06	.05	.03	.05
Originalitäts-Gesamtwert	.00	-.01	-.04	-.01

*) Eine Zufallswahrscheinlichkeit von $p \leq .05$ ist durch s, von $p \leq .01$ durch ss gekennzeichnet. (Ein scheinbarer Widerspruch kam durch Aufrundung zustande.)

Die für die Gesamtschüler ermittelten Korrelationskoeffizienten zeigen, daß die Kinder aus höheren Sozialschichten nur in 5 der 11 Aufgaben etwas bessere Werte erzielen. Die Zusammenhänge sind allerdings

1) Programm MIS 1 des Deutschen Rechenzentrums; Autor: L. Poecker, DRZ Darmstadt.

so gering, daß die Variablen des Sozialstatus auch bei Aufgabe 1, für die die höchsten Koeffizienten errechnet wurden, nur 3,6 Prozent der Varianz aufklären. Es erscheint fraglich, ob ein so geringfügiger schichtspezifischer Bias Anlaß zu Modifikationen geben sollte, zumal völlig offen ist, ob die Übereinstimmungs- und/oder Vorhersagevalidität dieser Aufgaben von denen der übrigen Aufgaben abweicht und wenn ja, in welcher Richtung.

Das für die Aufgabe 1 ungünstige Ergebnis gilt im übrigen nur für die Gesamtschulstichprobe; in den anderen 3 Schultypen bestehen keine signifikanten Zusammenhänge zwischen den Variablen des Sozialstatus und den Ergebnissen der Aufgabe 1. Während in Haupt- und Realschulen für 2 bzw. 3 Aufgaben einige signifikant positive Zusammenhänge gefunden wurden, ergaben sich für die 183 Gymnasiasten, für die diese Koeffizienten berechnet werden konnten, nur negative Zusammenhänge und zwar ausschließlich für die letzten 3 Aufgaben ("Bilderraten", also die 3 Aufgaben, die analog zu den "pattern meanings" von Wallach und Kogan konstruiert sind). In diesen Aufgaben erzielten diejenigen Kinder, deren Väter einen niedrigeren Schulabschluß haben, etwas bessere Ergebnisse; die höchsten Koeffizienten von -.27 und -.23 wurden für die Beziehung zwischen Schulbildung des Vaters und den beiden Ergebnissen der Aufgabe 1o berechnet. (Ob dieses Ergebnis auf die Gesamtheit der Gymnasien generalisiert werden kann, ist allerdings fraglich, da in dieser Stichprobe weder die übliche positive Korrelation der Intelligenztestergebnisse mit den Sozialschichtvariablen noch die ebenfalls oft festgestellte Beziehung zwischen Zensuren und Schicht-Indices gefunden wurde. Eventuell sind diese unerwarteten Ergebnisse durch die freiwillige Teilnahme der Schulen an der Untersuchung zu erklären.)

Für die Gesamtstichprobe aus allen 4 Schultypen liegen nur die Koeffizienten für die Zusammenhänge zwischen den Produktivitäts- und Originalitäts-Gesamtwerten und den 4 Statusmaßen vor. Tabelle 27 zeigt die Korrelationen.

Tabelle 77: Produkt-Moment-Korrelationen zwischen Produktivitäts- und Originalitäts-Gesamtwerten der Versuchsform VI, K und 4 Maßen des Sozialstatus für die Gesamtstichprobe.

	Vaterberuf N = 1o12	Schulbildung des Vaters N = 996	Schulbildung der Mutter N = 1oo5	Sozialstatus nach Bauer N = 1o14
Produktivitäts-Gesamtwert	r = .o7s*	r = .o43	r = .o93ss*	r = .o87ss
Originalitäts-Gesamtwert	r = .o38	r = .o17	r = .o64s	r = .o54

*) Eine Zufallswahrscheinlichkeit von $p \lesssim .o5$ ist durch s, von $p \lesssim .o1$ durch ss gekennzeichnet.

Die Gesamtwerte der Versuchsform VI, K werden in äußerst geringem Ausmaß durch die Variablen des Sozialstatus beeinflußt.
Dies gilt sehr wahrscheinlich auch für den Fall, daß eventuelle gegenläufige Tendenzen der Gymnasialstichprobe die Höhe der Zusammenhänge etwas verringert haben.
Den größten Effekt hat die Schulbildung der Mutter. Wenn man davon ausgeht, daß diese das für die Erziehung bedeutsamste Statusmaß ist - jedenfalls bedeutsamer als Beruf und Schulbildung des Vaters -, dann ist das ein plausibles Ergebnis.
Die Zusammenhänge der Statusvariablen mit dem Originalitäts-Gesamtwert sind niedriger und überwiegend nicht signifikant. Dies spricht nicht unbedingt dafür, daß dieses Maß des divergenten Denkens schichtunabhängiger ist, sondern ist eventuell allein auf die geringere Reliabilität der Originalitätswerte zurückzuführen.
Zusammenfassend kann festgestellt werden, daß weder Produktivitäts- noch Originalitätswert der Versuchsform VI, K mit dem Vaterberuf, der Schulbildung der Eltern oder einem kombinierten Maß des Sozialstatus substantiell korrelieren. In der Gesamtschulstichprobe sind sämtliche Korrelationen insignifikant, in der 4 Schularten umfassenden Gesamtstichprobe sind die wenigen signifikanten Korrelationen so niedrig, daß die durch die Sozialschicht-Variablen determinierte Varianz der Testergebnisse weniger als ein Prozent beträgt.

3.3314 Diagnostische Valenz der Originalitätswerte

Die letzte Frage, die durch die Untersuchungsergebnisse der Gesamtstichprobe beantwortet werden sollte, war diejenige nach dem Zusammenhang zwischen den Produktivitäts- und Originalitätsscores.[1)]

1) S. dazu S. 2oo-2o1

Tabelle 28 zeigt die Korrelationen der Anzahl- und Originalitätswerte der 1o Einzelaufgaben und die Korrelation der beiden Gesamtwerte. Die Koeffizienten beruhen auf den Testergebnissen der Gesamtstichprobe.

Tabelle 28: Produkt-Moment-Korrelationen zwischen Anzahl- und Originalitätsscores der 1o Aufgaben und der beiden Gesamtwerte der Versuchsform VI, K (N = 1671)

Aufgabe 2	.88
Aufgabe 3	.8o
Aufgabe 4	.81
Aufgabe 5	.8o
Aufgabe 6	.89
Aufgabe 7	.84
Aufgabe 8	.8o
Aufgabe 9	.9o
Aufgabe 1o	.83
Aufgabe 11	.92
Gesamtwerte	.92*

*) Der Koeffizient von .92 dürfte eine zuverlässige Schätzung des Zusammenhangs sein, nachdem die Korrelation der beiden Gesamtwerte in der aus 873 Gesamtschülern bestehenden Stichprobe fast ebenso hoch ist (r = .91).

Die Koeffizienten sind wesentlich höher als bei Versuchsform IV, 1, die Untersuchungshypothese wurde also bestätigt. Da wegen der geringen Aufgabenzahl der Versuchsform VI, K diagnostische oder prognostische Fragestellungen nur aufgrund der Gesamtwerte, nicht jedoch aufgrund der Ergebnisse einzelner Aufgaben oder Aufgabengruppen beantwortet werden können und da die Korrelation der beiden Gesamtwerte .92 beträgt, wird die Auswertung nach der Originalität der Antworten im Sinn ihrer statistischen Seltenheit als redundant betrachtet. Die Originalitätswerte werden daher bei der Darstellung der Untersuchungsergebnisse der Unterstichprobe nicht mehr berücksichtigt.
Die im Vergleich zu den in der Literatur berichteten Werten erheblich engeren Beziehungen zwischen Anzahl und Originalität[1)] werden auf die

1) Auch hier sind die Untersuchungsergebnisse nicht ganz widerspruchsfrei: Ward (1968) fand für 3 Subtests von Wallach und Kogan z.T. noch höhere Korrelationen und Grote, Hajek u.a. (1969) erhielten für einige Items ebenfalls sehr hohe Koeffizienten (vgl. Tabelle 5, S.13o).

erhöhte Varianz der Testergebnisse und auf die Einführung eines qualitativen Auswertungskriteriums für die Produktivitätsscores zurückgeführt.[1] Welcher der beiden Faktoren den stärkeren Einfluß hatte, kann nicht beurteilt werden.

3.332 Ergebnisse der Teilstichprobe

An der 5 Klassen aus 3 Gesamtschulen umfassenden Teilstichprobe[2] wurden zwei Untersuchungen zur Übereinstimmungsvalidität und zwei Überprüfungen der Konstruktvalidität des Verfahrens durchgeführt. Vor der Darstellung dieser Ergebnisse werden einige Merkmale der untersuchten Stichprobe besprochen.

3.3321 Stichprobenmerkmale

Da es den Klassenlehrern freigestellt war, ob sie mit ihren Schülern an diesen Untersuchungen teilnehmen oder nicht, wird zunächst überprüft, ob sich die für diese 5 Klassen berechneten Mittelwerte und Streuungen der Produktivitäts-Gesamtwerte der Versuchsform VI, K von den entsprechenden Ergebnissen der Gesamtschulstichprobe unterscheiden.
Tabelle 29 ermöglicht einen Vergleich der Resultate beider Stichproben. Die Signifikanzprüfung wurde mit Hilfe der Standardfehler des arithmetischen Mittels und der Standardabweichung vorgenommen, wobei die Gesamtschulstichprobe als Population betrachtet wurde, aus der die Unterstichprobe als Zufallsstichprobe gezogen worden war. Bei Abweichungen zwischen den Mittelwerten und Streuungen, die mit einer Zufallswahrscheinlichkeit von 5 Prozent oder weniger zu erwarten sind, sollte die Hypothese, daß die Unterstichprobe eine Zufallsstichprobe aus allen untersuchten Gesamtschulklassen ist, zurückgewiesen werden.

Tabelle 29: Arithmetische Mittel ($\bar{X}$) und Standardabweichungen (σ) der Produktivitäts-Gesamtwerte der Gesamtschul- und der daraus entnommenen Unterstichprobe.

	Unterstichprobe (N = 137)	Gesamtschulstichprobe (N = 873)	Signifikanz der Abweichung
$\bar{X}$	57,94	55,7o	nicht signifikant
σ	17,49	17,52	nicht signifikant

1) Vgl. S. 2oo-2o1

2) Zur Beschreibung der Teilstichprobe und der zusätzlich durchgeführten Verfahren s.S. 197

Weder die Mittelwerte noch die Standardabweichungen unterscheiden sich statistisch bedeutsam.

Da die Resultate des Persönlichkeitsfragebogens CPQ für Jungen und Mädchen getrennt verrechnet werden sollten, mußte auch untersucht werden, ob die geschlechtsspezifischen Unterschiede in den Ergebnissen der Versuchsform VI, K, die für die Gesamtschüler gefunden wurden, mit denjenigen der Teilstichprobe übereinstimmen. Eine Überprüfung der Testergebnisse der 7o Jungen und 66 Mädchen[1] erbrachte in der Tendenz ebenfalls ähnliche Resultate wie für die Gesamtschulstichprobe.

Tabelle 3o: Arithmetische Mittel der Produktivitätswerte der 11 Aufgaben und des Produktivitäts-Gesamtwerts der Versuchsform VI, K für die Jungen und Mädchen der Unterstichprobe sowie t-Werte des Mittelwertsvergleichs.

Aufgabe	$\bar{X}$ Jungen	$\bar{X}$ Mädchen	t
1	6,81	5,5o	2,11 s*
2	4,41	4,74	o,99
3	3,86	4,11	o,79
4	5,54	5,53	o,12
5	4,41	5,17	2,22 s
6	4,1o	4,86	2,2o s
7	4,59	4,8o	o,64
8	5,21	5,65	1,18
9	6,61	8,17	2,42 s
1o	4,31	5,o9	1,82
11	5,83	6,68	1,41
Gesamtwert	55,7o	6o,35	1,55

*) s bedeutet $p \leq .o5$

In der ersten Aufgabe schneiden die Jungen beider Stichproben besser ab, die Differenz ist jedoch nur in der Unterstichprobe signifikant.[2] Alle anderen signifikanten Mittelwertsdifferenzen sind zugunsten der Mädchen. Da in der großen Gesamtschulstichprobe die gleichen Aufgaben ebenfalls von den Mädchen signifikant besser beantwortet wurden und hier nur einige weitere Signifikanzen hinzukommen, führt

1) Ein Pb konnte wegen fehlender Angaben nicht zugeordnet werden.

2) Vgl. S. 222

auch die Überprüfung der geschlechtsspezifischen Differenzen zu dem Ergebnis, daß die 5 Klassen der Unterstichprobe hinsichtlich der Ergebnisse der Versuchsform VI, K als Zufallsstichprobe aus der 34 Klassen umfassenden "Gesamtschulpopulation" betrachtet werden können.

Ein Vergleich der klassenspezifischen Mittelwerte der Produktivitäts-Gesamtsscores zeigt dagegen, daß zwischen den 5 Klassen erhebliche Unterschiede bestehen (vgl. Tabelle 31).

Tabelle 31: Mittelwerte und Standardabweichungen der Produktivitäts-Gesamtwerte für die 5 Klassen der Unterstichprobe

	1 N = 30	2 N = 29	3 N = 30	4 N = 23	5 N = 25
$\bar{x}$	53,07	45,93	60,57	65,87	67,28
σ	13,31	14,39	15,22	19,04	17,39

Während die Überprüfung der Homogenität der Varianzen nur für die Klassen 1 und 4 eine auf dem 5 %-Niveau signifikante Differenz erbrachte,[1)] zeigen die Klassenmittelwerte für die überwiegende Zahl der möglichen Kombinationen statistisch bedeutsame Unterschiede. Tabelle 32 enthält die Ergebnisse des t-Tests zur Überprüfung der Mittelwertsdifferenzen.

Tabelle 32: Wahrscheinlichkeiten (p) für die Ergebnisse des t-Tests zur Überprüfung der Mittelwertsdifferenzen der Produktivitäts-Gesamtwerte aus Tabelle 31

Klassen	Signifikanz der t-Werte
1/2	n. s.*
1/3	$p < .05$
1/4	$p < .01$
1/5	$p < .001$
2/3	$p < .001$
2/4	$p < .001$
2/5	$p < .001$
3/4	n. s.
3/5	n. s.
4/5	n. s.

*) n. s. = nicht signifikant ($p > .05$)

1) Prüfung mit dem F-Test nach R. A. Fischer

Daß sich die Mittelwerte verschiedener Klassen erheblich unterscheiden, war aufgrund der Ergebnisse von Krause (1972) erwartet worden.[1] Die Bestätigung dieser Erwartung kompliziert die Analyse der an der Unterstichprobe erhobenen Daten, weil in den meisten Fällen geprüft werden muß, ob die Resultate der Unterstichprobe auch für die einzelnen Klassen Gültigkeit besitzen.

3.3322 Ergebnisse der Untersuchung mit dem Lehrerfragebogen

In 4 fünften Klassen einer Bad Dürkheimer Hauptschule hatte die Überprüfung der Übereinstimmungsvalidität der Versuchsform IV, 1 mit dem Lehrerfragebogen keine befriedigenden Ergebnisse erbracht. Da nicht auszuschließen war, daß dies auf Besonderheiten der Stichprobe zurückzuführen ist, wurde diese Untersuchung - nur geringfügig modifiziert[2] - mit der Versuchsform VI, K wiederholt.

Eine faktorenanalytische Überprüfung der von 13 auf 9 Skalen reduzierten Version des Fragebogens[3] ergab, daß die Beurteilungsstruktur der Lehrer der 5 Gesamtschulklassen mit derjenigen der 4 Hauptschullehrer aus Bad Dürkheim weitestgehend übereinstimmt.[4] Bei einer Eigenwertgrenze von o.5 ergaben sich wieder nur 2 Faktoren, die zusammen 74,1 Prozent der Gesamtvarianz aufklären. Auf dem ersten Faktor, der mit 69 Prozent an der Varianzaufklärung beteiligt ist, laden die Fragen 2, 4, 5, 7, 8 und 9 mit Ladungen zwischen .74 und .95. Der zweite Faktor ist mit 32 Prozent an der Varianzaufklärung beteiligt und wird durch die Fragen 1, 3 und 6 repräsentiert; die Ladungen liegen zwischen .74 und .83. Faktorenanalysen mit den Daten jeder einzelnen der 5 Klassen erbrachten keine deutlichen Abweichungen von dieser Faktorenstruktur.

Somit war es möglich, die Ratings zu den Fragen 2, 4, 5, 7, 8 und 9 zu dem Gesamtwert "Motivation für kreatives Verhalten" und die Einstufungen zu den Fragen 1, 3 und 6 zu dem Gesamtwert "Anpassung an die Schulsituation" zusammenzufassen.

Wie in der Bad Dürkheimer Stichprobe ergab sich auch in den 5 Gesamtschulklassen zwischen dem ersten Gesamtwert des Lehrerfragebogens und dem Produktivitäts-Gesamtwert der Versuchsform VI, K eine wesentlich niedrigere Korrelation als zwischen dem Lehrerurteil und einer Reihe schulisch relevanter Variablen. Tabelle 33 zeigt die Korrelationen zwischen dem ersten Gesamtwert der Lehrerfragen, Kreativitäts- und Intelligenztestergebnissen, Zensuren und den Schülerratings.

---

1) Vgl. S. 1o7-1o8

2) S. dazu S. 192

3) Die reduzierte Version des Fragebogens findet sich in Anhang A 3. Zu seiner Entwicklung s.S. 165-171, Hypothesen zu seinem Einsatz in der Hauptuntersuchung s.S. 2o1.

4) Programm PAFA des DRZ; s. dazu Anmerkung 1), S. 156

Tabelle 33: Produkt-Moment-Korrelationen zwischen dem ersten Gesamtwert der Lehrerfragen ("Motivation für kreatives Verhalten") und ausgewählten Schülervariablen

Schülervariablen	N	1. Gesamtwert der Lehrerfragen
Geschlecht* (männlich = 1, weiblich = 2)	136	- o.247 ss**
Versuchsform VI, K	137	o.227 ss
Kreativitätstest von Wallach und Kogan	118	o.174 s**
P-S-B	117	o.5o6 ss
BBT	12o	o.645 ss
Zensur in Deutsch	131	- o.55o ss
" " Mathematik	131	- o.592 ss
" " 1. Fremdsprache	131	- o.476 ss
" " Kunst	131	- o.342 ss
Beruf des Vaters	99	o.254 ss
Schulbildung des Vaters	97	o.3o3 ss
" der Mutter	99	o.384 ss
Sozialstatus nach Bauer	99	o.353 ss
Lehrerurteil über Fleiß***	11o	- o.467 ss
" " Mitarbeit	131	- o.632 ss
" " Ehrgeiz	111	- o.395 ss
" " Ausdauer	13o	- o.46o ss
1. Schülerfrage	136	o.644 ss
2. "	136	o.715 ss
3. "	136	o.667 ss
Gesamtwert der Schülerfragen	136	o.714 ss

*) Zur Berechnung dieses Korrelationskoeffizienten s. Anmerkung 1), S. 172

**) ss bedeutet $p \geq .01$, s bedeutet $.10 > p > .05$

***) Analog zur Notenskala, Stufen von 1 - 6

Der erste Gesamtwert der Lehrerfragen korreliert in dieser Stichprobe wesentlich höher mit den Schülervariablen als in der Bad Dürkheimer Stichprobe. Dies ist nicht auf eine höhere Streuung der Lehrerurteile zurückzuführen: der Variabilitätskoeffizient des aus 6 Ratings gebildeten ersten Gesamtwerts beträgt in der aus 5 Gesamtschulklassen bestehenden Unterstichprobe o.276; in der Bad Dürkheimer Stichprobe beträgt dieses Streuungsmaß des aus 7 Ratings gebildeten Gesamtwerts o.311.

Die in Tabelle 33 dargestellten Zusammenhänge lassen vermuten, daß das Lehrerurteil über die Motivation der Schüler zu kreativem Verhalten stark an der Intelligenz und den Schulleistungen orientiert ist und daß Jungen sowie Kinder aus Familien mit höherem Bildungsniveau höher eingeschätzt werden. Die Korrelationen zwischen dem Lehrerurteil und den Schülerratings sind angesichts der Tatsache, daß es sich um 5 Klassen aus 3 verschiedenen Schulen handelt, auffallend hoch.

Mit der Versuchsform VI, K korreliert der erste Gesamtwert der Lehrerfragen in der gesamten Unterstichprobe zwar hochsignifikant, aber niedrig.

Diese Beziehung ist zwar von Klasse zu Klasse recht unterschiedlich, aber in keiner Klasse sehr hoch ($.025 \leq r \leq .406$; $\bar{r} = .27$). Dasselbe gilt für die Beziehung zwischen dem Lehrerurteil und den Ergebnissen des Kreativitätstests von Wallach und Kogan ($-.252 \leq r \leq .454$; $\bar{r} = .18$); die über z-Transformation gemittelten Korrelationskoeffizienten der 5 Klassen sind nur unerheblich höher als die für alle Schüler der Teilstichprobe berechneten Koeffizienten.

Die von Klasse zu Klasse unterschiedlichen Varianzen des Lehrerurteils bieten keine ausreichende Erklärung für die differierenden Korrelationen. Während die Zusammenhänge zwischen den 9 Skalen des Lehrerfragebogens in den 5 Klassen sehr gut übereinstimmen, sind die Beziehungen des aus 6 Skalen gebildeten Gesamtwertes "Motivation für kreatives Verhalten" zu verschiedenen Untersuchungsvariablen von Klasse zu Klasse sehr unterschiedlich. Tabelle 34 zeigt für jede Klasse die 2 Variablen, die mit dem Lehrerurteil über die Motivation zu kreativem Verhalten am höchsten korrelieren.

Tabelle 34: Übersicht über die mit dem ersten Gesamtwert der Lehrerfragen am höchsten korrelierenden Variablen für die 5 Klassen der Unterstichprobe

Klasse	Variable	r	N	Variable	r	N
1	Untertest Wortbedeutungen des BBT	.745	25	Untertest Satzteile des BBT	.732	25
2	Zensur in Mathematik	-.802	29	Untertest 4 (Buchstabenreihen)des I-S-B	.788	19
3	Lehrerurteil über Mitarbeit*	-.802	28	Lehrerurteil über Fleiß*	-.800	28
4	Lehrerurteil über Mitarbeit*	-.673	19	Item 31 des Kreativitätstests von Wallach und Kogan	.576	19
5	Gesamtwert des BBT	.721	24	Zensur in Kunst	-.687	25

*) Analog zur Notenskala, Stufen 1 - 6

Bereits diese kleinen Ausschnitte aus den 5 Korrelationsmatrizen zeigen, daß die Ratings der 5 Lehrer über die Motivation der Schüler zu kreativem Verhalten von sehr verschiedenen Beurteilungsgrundlagen ausgehen. Hinzu kommt, daß die Lehrerurteile in unterschiedlicher Weise durch das Geschlecht der Schüler beeinflußt werden: Die Korrelationen des ersten Gesamtwertes der Lehrerratings mit dem Geschlecht der Schüler liegen zwischen -o.66 und +o.14 (Md = -o.18); d.h. es gibt Klassen, in denen die Jungen besser beurteilt wurden als die Mädchen, und andere, in denen das Geschlecht für die Ratings keine Rolle spielte. (Für die gesamte Unterstichprobe ergibt sich eine signifikant bessere Beurteilung der Jungen: $t = 3.04$, $p < .01$).

Für den zweiten Gesamtwert der Lehrerfragen "Anpassung an die Erfordernisse der Schulsituation" war eine niedrigere Korrelation mit der Versuchsform VI, K erwartet worden als für den ersten Gesamtwert. Die Zusammenhänge zwischen diesem Ergebnis des Lehrerfragebogens und den ausgewählten Schülervariablen werden in Tabelle 35 dargestellt.

Tabelle 35: Produkt-Moment-Korrelationen zwischen dem zweiten Gesamtwert der Lehrerfragen und ausgewählten Schülervariablen.

Schülervariablen	N	2. Gesamtwert der Lehrerfragen
Geschlecht* (männlich = 1, weiblich = 2)	136	o.322 ss**
Versuchsform VI, K	137	o.o66
Kreativitätstest von Wallach und Kogan	118	o.171
P-S-B	117	o.154
BBT	12o	o.159
Zensur in Deutsch	131	o.262 ss
" " Mathematik	131	o.195 s**
" " 1. Fremdsprache	131	o.416 ss
" " Kunst	131	o.224 ss
Beruf des Vaters	99	o.268 ss
Schulbildung des Vaters	97	o.2o3 s
" der Mutter	99	o.197 s
Sozialstatus nach Bauer	99	o.274 ss
Lehrerurteil über Fleiß***	11o	o.717 ss
" " Mitarbeit	131	o.385 ss
" " Ehrgeiz	111	o.547 ss
" " Ausdauer	13o	o.65o ss
1. Schülerfrage	136	o.43o ss
2. Schülerfrage	136	o.391 ss
3. Schülerfrage	136	o.oo7
Gesamtwert der Schülerfragen	136	o.287 ss
1. Gesamtwert der Lehrerfragen	137	o.124

*) Zur Berechnung dieses Koeffizienten s. Anmerkung 1), S. 172

**) ss bedeutet $p \leq .01$, s bedeutet $p \leq .05$.

***) Analog zur Notenskala, Stufen von 1 - 6

Die Korrelationen der Tabelle 35 zeigen, daß das Lehrerurteil über die Anpassung der Schüler an die Schulsituation eng mit ihrer Beurteilung des Fleißes und der Ausdauer zusammenhängt. Eine Überprüfung der Korrelationsmatrizen der 5 Klassen ergibt, daß in allen Klassen die Lehrerurteile über Fleiß und Ausdauer oder Ehrgeiz am höchsten mit dem zweiten Gesamtwert der Lehrerfragen korrelieren.

Diese Variablen wirken sich offenbar auch auf die Leistungen in der ersten Fremdsprache aus.
Bei dem zweiten Gesamtwert der Lehrerfragen schneiden die Mädchen und - wie beim ersten Gesamtwert - die Kinder aus Familien mit höherem Sozialstatus besser ab. (Die geschlechtsspezifischen Mittelwerte dieses Lehrerurteils unterscheiden sich wieder hochsignifikant: $t = 3{,}62$, $p < .001$)
Die Schülerurteile über das Verhalten in kreativitätsfordernden Situationen werden durch die schulische Anpassung wesentlich weniger stark beeinflußt als durch die Verhaltensweisen, deren Beurteilung zum ersten Gesamtwert der Lehrerfragen zusammengefaßt wurden (vgl. Tabelle 33 und 35). (In der Bad Dürkheimer Stichprobe waren dagegen die Zusammenhänge zwischen den Schülerurteilen und den beiden Gesamtwerten der Lehrerfragen etwa von gleicher Höhe.[1])
Sämtliche angesprochenen Beziehungen sprechen dafür, daß der zweite Gesamtwert der Lehrerfragen tatsächlich den Grad der Angepaßtheit an die derzeitige Schulsituation wiedergibt.
Die Untersuchungshypothese, daß dieser Wert mit der Versuchsform VI,K weniger eng zusammenhängt als das Lehrerurteil über die Motivation zu kreativem Verhalten, wurde bestätigt: die Korrelation liegt bei Null.[2]

Eine zusammenfassende Beurteilung der Ergebnisse der Untersuchung mit dem Lehrerfragebogen kann der Bestätigung dieser zweiten Hypothese nicht allzu viel Bedeutung beimessen: wichtigste Resultate sind die zwar hochsignifikante, aber niedrige Korrelation des Lehrerurteils über die Motivation der Schüler zu kreativem Verhalten mit den Ergebnissen der Versuchsform VI, K und die wesentlich engeren Zusammenhänge zwischen diesem Lehrerurteil und Schulleistungs- und Intelligenzvariablen. Vor allem die hohen Korrelationen mit 2 Intelligenztests stellen die Validität des Außenkriteriums in Frage.

Mit Sicherheit kann allerdings nur gesagt werden, daß die Ergebnisse der Untersuchung mit dem Lehrerfragebogen keine Aufschlüsse über die Übereinstimmungsvalidität der Versuchsform VI, K erbracht haben.

---

1) S. dazu Tabelle 13 und 14, S.172 und S. 174

2) In der heterogeneren Stichprobe der Hauptuntersuchung entsprechen somit die Ergebnisse eher den Hypothesen als in der aus 4 Klassen einer Hauptschule bestehenden Stichprobe der Voruntersuchung (s. S. 171-176).

3.3323 Ergebnisse der Untersuchung mit den Schülerfragen[1)]

Wie in den 4 Hauptschulklassen interkorrelieren die Ergebnisse der 3 Schülerfragen auch in den 5 Gesamtschulklassen sehr hoch ($.72 \leq r \leq .90$).

Die Modifikation der 3. Frage "lustige Streiche", die den geschlechtsspezifischen Bias dieser Frage eliminieren sollte, hatte nicht den gewünschten Effekt: die Korrelation der Schülerratings mit dem Geschlecht beträgt -o.34, d.h. die Fähigkeit der Jungen im Ausdenken "lustiger Überraschungen" wird höher eingeschätzt als diejenige der Mädchen.

Da die Streuung der Versuchsform VI, K mit δ = 17.49 größer ist als diejenige der Versuchsform IV, 1, für die die Standardabweichung 12.63 betrug, wurden engere Zusammenhänge mit den Ergebnissen der Schülerfragen erwartet. Das Gegenteil ist der Fall.

Tabelle 36 zeigt die Korrelationen des aus den Ergebnissen der 3 Ratings gebildeten Gesamtwerts der Schülerbefragung mit den ausgewählten Schülervariablen.

Tabelle 36: Produkt-Moment-Korrelationen zwischen dem Gesamtwert der Schülerfragen und ausgewählten Variablen.

Ausgewählte Variablen	N	Gesamtwert der Schülerfragen
Geschlecht* (männlich = 1, weiblich = 2)	135	- o.1o6
Versuchsform VI, K	136	o.169 s**
Kreativitätstest von Wallach und Kogan	117	o.358 ss**
P-S-B	116	o.338 ss
BBT	119	o.45o ss
Zensur in Deutsch	13o	- o.376 ss
" " Mathematik	13o	- o.4oo ss
" " 1. Fremdsprache	13o	- o.472 ss
" " Kunst	13o	- o.268 ss
Beruf des Vaters	98	- o.o29
Schulbildung des Vaters	96	o.o92
" der Mutter	98	o.198 s
Sozialstatus nach Bauer	98	o.o75
Lehrerurteil über Fleiß***	1o9	- o.471 ss
" " Mitarbeit	13o	- o.474 ss
" " Ehrgeiz	11o	- o.382 ss
" " Ausdauer	12o	- o.5oo ss

*) Zur Berechnung dieses Koeffizienten s. Anmerkung 1), S.172

**) s bedeutet $p \leq .05$, ss bedeutet $p \leq .01$

***) Analog zur Notenskala, Stufen von 1 - 6

1) Zur Entwicklung der Schülerfragen s.S.178-181, zu Modifikationen s. S.192-193 und Anmerkungen zu den Schülerfragen s.S.2o2

Die Korrelationen des Gesamtwerts der Schülerfragen sind insgesamt nicht sehr hoch, der höchste Koeffizient wurde für die Beziehung zu dem Lehrerurteil über die Ausdauer der Schüler errechnet. Zu den Lehrerurteilen über Fleiß, Mitarbeit und Ehrgeiz, zu den Noten der 3 Hauptfächer sowie zu den beiden Intelligenztestergebnissen bestehen mäßige Korrelationen.

Unerwartet niedrig ist der Zusammenhang zwischen dem Gesamtwert der Schülerratings und den Ergebnissen der Versuchsform VI, K. Er ist nur auf dem 5-Prozent-Niveau signifikant und ist sogar noch geringer als die Korrelation mit der Schulbildung der Mutter.

Im Vergleich dazu korreliert das Ergebnis des Kreativitätstests von Wallach und Kogan wesentlich höher mit den Schülerurteilen, was insofern überrascht, als dieses Verfahren in früheren Untersuchungen mit Lehrer- und Schülerratings weniger hoch korrelierte als die Vorformen von VI, K.[1] Der Grund für diese relativ enge Beziehung ist eventuell darin zu sehen, daß die Schülerratings durch Verhaltensweisen beeinflußt wurden, die das Lehrerurteil über die Ausdauer der Schüler determinierten (der Gesamtwert der Schülerfragen korreliert .5o mit dem Lehrerurteil über die Ausdauer der Schüler, vgl. Tabelle 36). Da ein hohes Maß an Ausdauer auch bei der Beantwortung der 39 Items des Kreativitätstests von Wallach und Kogan von Vorteil ist, wird vermutet, daß in der von Lehrer und Schülern wahrgenommenen Persistenz des Verhaltens der Grund für die vergleichsweise enge Beziehung zwischen dem Gesamtwert der Schülerfragen und dem Ergebnis des Kreativitätstests von Wallach und Kogan liegt. Die Frage, warum die Ausdauer der Schüler die Schülerratings so stark beeinflußte, muß allerdings offen bleiben.

Der Zusammenhang zwischen den Ergebnissen der Versuchsform VI, K und dem Gesamtwert der Schülerfragen wurde auch klassenweise untersucht. Die Korrelationen in den 5 Klassen zeigen wieder stärkere Abweichungen ($.o8 \leq r \leq .47$). Die über z-Transformation berechnete mittlere Korrelation beträgt nur .34.

Die höchsten klassenspezifischen Korrelationen wurden für den Zusammenhang zwischen der Versuchsform VI, K und der ersten der 3 Schülerfragen ("Theaterstück ausdenken") errechnet ($.17 \leq r \leq .57$, $\bar{r} = .38$).

Die Korrelation aller 3 Fragen und des Gesamtwerts mit den Ergebnissen der Versuchsform VI, K sind zwar hochsignifikant, aber in Anbetracht

---

1) Vgl. Tabelle 13, S. 172, und Tabelle 17, S. 182

der hohen Reliabilität der aus einer großen Zahl von Schätzurteilen berechneten Ratings sehr niedrig.
Eine Inspektion der Korrelationsmatrizen im Hinblick auf Beziehungen der Ratings zu anderen Variablen erbringt wieder für jede der 5 Klassen ein anderes Bild. Tabelle 37 zeigt für jede Klasse die beiden Variablen, mit denen der Gesamtwert der Schülerfragen am höchsten korreliert.

Tabelle 37: Übersicht über die mit dem Gesamtwert der Schülerfragen am höchsten korrelierenden Variablen für die 5 Klassen der Unterstichprobe

Klasse	Variable	r	N	Variable	r	N
1	Lehrerurteil über Mitarbeit*)	-o.74	3o	Zensur in Deutsch	-o.66	3o
2	Untertest Denkaufgaben des BBT	o.74	27	Zensur in Mathematik	-o.7o	29
3	Lehrerurteil über die Ausdauer*	-o.83	27	Lehrerurteil über Ehrgeiz*	-o.8o	27
4	Aufgabe 27 des Tests von Wallach und Kogan	o.61	19	Aufgabe 35 des Tests von Wallach und Kogan	o.59	19
5	Skala "Prüfungsangst" des AFS	-o.63	25	BBT-Gesamtwert	o.61	24

*) Analog zur Notenskala, Stufen von 1 - 6

Die hochsignifikanten und zum überwiegenden Teil sehr hohen Korrelationen mit völlig verschiedenen Variablen lassen vermuten, daß der Gesamtwert der Schülerfragen in allen Klassen aufgrund unterschiedlicher Beurteilungskriterien zustande kam. Klassenübergreifende Berechnungen können daher gar keine sehr hohen Korrelationskoeffizienten erbringen. Im übrigen sind sie in Anbetracht der von Klasse zu Klasse sehr unterschiedlichen Zusammenhänge methodisch angreifbar, da sie zu irreführenden Aussagen führen.
Aber selbst eine klassenspezifische Verarbeitung der Untersuchungsdaten erscheint angesichts der stark abweichenden Korrelationen des Gesamtwerts der Schülerfragen mit dem Geschlecht der Probanden problematisch ($-$ o.487 $\leq r \leq +$ o.381, Md = $-$ o.o84). Für die gesamte Unterstichprobe ergeben sich - wie nach dem Median der Korrelationen zu vermuten - keine Unterschiede zwischen dem Mittelwert der Schülerfragen der Mädchen und demjenigen der Jungen ($t = 1.34$, $p > .1o$).

Die Gegenüberstellung der klassenspezifisch und klassenübergreifend berechneten Zusammenhänge stellt auch Ergebnisse der Untersuchung mit dem Lehrerfragebogen in der Bad Dürkheimer Stichprobe in Frage. Vor allem Aussagen über fehlende Zusammenhänge müssen nicht für alle 5 Klassen zutreffen. So besteht beispielsweise die Möglichkeit, daß in einem Teil der Klassen eine signifikant negative und in dem anderen Teil eine signifikant positive Beziehung zwischen dem ersten Gesamtwert der Lehrerfragen und den Kreativitätstestergebnissen besteht.[1)]

Somit war der Versuch, aufgrund von Schülerurteilen die Übereinstimmungsvalidität der Versuchsform VI, K nachzuweisen, auch in einer heterogenen Schülerstichprobe nicht erfolgreich.

Die Resultate der Schülerbefragung lassen ebenso wie diejenigen der Untersuchung mit dem Lehrerfragebogen die Frage der Gültigkeit der Versuchsform VI, K unbeantwortet. Zugleich stellen sie die Aussagekraft aller Untersuchungen in Frage, in denen zum Nachweis der Übereinstimmungsvalidität eines Kreativitätstests niedrige bis mäßige Korrelationen mit klassenweise erhobenen Befragungsergebnissen mitgeteilt werden.

3.3324 Ergebnisse der Untersuchung mit dem Test von Wallach und Kogan und zwei Intelligenztests

Wenngleich die Ergebnisse der Untersuchung mit dem Kreativitätstest von Wallach und Kogan auch eine Beurteilung der Übereinstimmungsvalidität der Versuchsform VI, K ermöglichen, wurde der im folgenden dargestellte Teil der Arbeit in erster Linie mit dem Ziel der Konstruktvalidierung des eigenen Verfahrens durchgeführt.
Ein Aspekt der Konstruktvalidität der Versuchsform VI, K, die diskriminante Validität gegenüber Intelligenztestergebnissen, war mit gutem Resultat an der Gesamtstichprobe überprüft worden.[2)] Die Hypothese, daß das Verfahren Faktoren erfaßt, die auch durch die Subtests der traditionellen Intelligenztests gemessen werden, kann aufgrund der mit der Versuchsform VI, K und einer größeren Zahl von Intelligenztest- und Schulleistungsvariablen durchgeführten Faktorenanalysen als widerlegt betrachtet werden.[3)]

---

1) Vgl. Tabelle 13, S. 172

2) S. dazu S.213-221

3) Auch die mit der Versuchsform IV, 1 und anderen Intelligenztest- und Schulleistungsvariablen durchgeführte Faktorenanalyse führt zu demselben Schluß, vgl. S. 155-160.

Damit ist natürlich nur eine von vielen möglichen Alternativhypothesen über das durch das Verfahren gemessene Konstrukt überprüft. Nach Cronbach (1971) stehen die Aufstellung und empirische Prüfung solcher Alternativhypothesen im Mittelpunkt des langwierigen Prozesses der Konstruktvalidierung.
Allerdings werden die Alternativen meist nicht von dem Testkonstrukteur, sondern von anderen Kennern des jeweiligen Gebiets vorgeschlagen[1], was in Anbetracht der Befangenheit des ersteren in seinen theoretischen Vorstellungen mit Sicherheit das effektivere Verfahren ist.

Deshalb sollen im Rahmen der Konstruktvalidierung der Versuchsform VI,K keine weiteren Untersuchungen zur diskriminanten Validität durchgeführt werden, sondern die Abgrenzung gegen die Testintelligenz als dem wichtigsten alternativen Konstrukt soll ergänzt werden durch die Überprüfung der konvergenten Validität von VI, K mit Hilfe des Kreativitätstests von Wallach und Kogan. Da für die Unterstichprobe der Hauptuntersuchung die Ergebnisse zweier traditioneller Intelligenztests und die Gesamtwerte zweier Kreativitätstests vorliegen, ist es möglich, die Resultate aller Verfahren in Form einer "multitrait-multimethod matrix"[2] darzustellen, in der die Koeffizienten der diskriminanten und der konvergenten Validität enthalten sind. Da für jedes der beiden Konstrukte - Testintelligenz und divergentes Denken - nur je 2 Meßverfahren eingesetzt wurden und da keine vergleichbaren Reliabilitätsmaße für die 4 Tests vorliegen, handelt es sich allerdings um eine etwas reduzierte multitrait-multimethod matrix.
Obwohl in diese Untersuchung weder Lehrer- noch Schülerurteile einbezogen sind, wurden die Koeffizienten sowohl für die gesamte Unterstichprobe als auch als Mittelwerte der klassenspezifischen Korrelationen berechnet. Nachdem beide Kreativitätstests im Klassenrahmen und mit den Klassenlehrern als Testleiter durchgeführt worden waren, wurden von den klassenspezifischen Ergebnissen Aufschlüsse über die Abhängigkeit der Gesamtwerte der beiden Verfahren von der jeweiligen Testsituation erwartet.

---

1) Z.B. stellt Guilford (1971) die Hypothese auf, daß der Kreativitätstest von Wallach und Kogan Persistenz oder Widerstand gegen Langeweile mißt und Cronbach (1968) überprüft die Alternativhypothese, daß dieses Verfahren "responsiveness" in spielähnlichem Kontext untersucht.

2) Campbell und Fiske 1959. In der Bad Dürkheimer Stichprobe waren zwar auch 2 Verfahren zur Erfassung des divergenten Denkens, aber nur ein Intelligenztest durchgeführt worden (vgl. Tabelle 19, S. 189).

Tabelle 38 zeigt die verschiedenen Korrelationskoeffizienten.

Tabelle 38: Korrelationen zweier Verfahren zur Erfassung des divergenten Denkens und zweier Intelligenztests. Die Koeffizienten der Unterstichprobe (r) stehen unterhalb, die gemittelten Koeffizienten der 5 Klasen ($\bar{r}$) oberhalb der Diagonale.

	Versuchsform VI, K	Kreativitätstest von Wallach und Kogan	P-S-B	BBT
Versuchsform VI, K		$\bar{r}$ = .65 ss**	$\bar{r}$ = .45 ss	$\bar{r}$ = .28 ss
Kreativitätstest von Wallach und Kogan	r = .46 ss N = 118*		$\bar{r}$ = .3o ss	$\bar{r}$ = .18
P-S-B	r = .41 ss N = 117	r = .21 s N = 1o6		$\bar{r}$ = .69 ss
BBT	r = .33 ss N = 12o	r = .23 s N = 1o4	r = .64 ss N = 1o7	

*) Dasselbe N gilt für den entsprechenden Koeffizienten oberhalb der Diagonale

**) ss bedeutet $p \leq .01$, s bedeutet $p \leq .05$

Eine Überprüfung mit dem t-Test von Hotelling[1] erbrachte für die gemittelten Korrelationskoeffizienten (oberhalb der Diagonale) das erwartete Resultat: der für die konvergente Validität errechnete Koeffizient von .65 ist signifikant höher als die Korrelationen der Versuchsform VI, K mit den beiden Intelligenztests von .45 und .28 ($t = 2{,}41$, $p < .02$, und $t = 3.91$, $p < .001$).

Die klassenübergreifend berechneten Korrelationen der Versuchsform VI, K mit den 3 anderen Meßverfahren (unterhalb der Diagonale) unterscheiden sich nicht signifikant.

Bei beiden Arten der Korrelationsberechnung sind die Zusammenhänge zwischen dem Kreativitätstest von Wallach und Kogan und den beiden Intelligenztests geringer als diejenigen zwischen der Versuchsform VI, K und den Intelligenztests. Für dieses Verfahren ergeben sich auch bei klassenübergreifender Berechnung der Korrelationen (unterhalb der Diagonale) signifikante Unterschiede zwischen dem Koeffizienten für die konvergente und den beiden Koeffizienten für die diskriminante Validität.

1) T-Test für abhängige Stichproben von Hotelling 194o, dargestellt in Edwards 1971, S. 97-98

Da mit der Versuchsform VI, K nicht Kreativität, sondern nur die Fähigkeit zum divergenten Denken gemessen werden soll, ist ihre etwas engere Beziehung zu Intelligenztestergebnissen nicht überraschend. Der Kreativitätstest von Wallach und Kogan dürfte allein aufgrund seiner hohen Itemzahl stärker von motivationalen Bedingungen abhängig sein als die Versuchsform VI, K.
Die gemittelten und die klassenübergreifend berechneten Korrelationskoeffizienten stimmen meist recht gut überein (vgl. die entsprechenden Koeffizienten rechts und links der Diagonalen in Tabelle 38), nur für die Beziehung zwischen den beiden Kreativitätstests ergeben sich mit .46 und .65 stärker differierende Werte. Dies ist vermutlich ein Hinweis auf die höhere Beeinflußbarkeit eines der beiden Tests oder beider Verfahren durch Änderungen der Testsituation.
Zur Überprüfung dieser Hypothese wurden die Ergebnisse der Versuchsform VI, K für die 5 Klassen mit denjenigen des Kreativitätstests von Wallach und Kogan verglichen. Es ergaben sich einige starke Abweichungen in den aufgrund der Mittelwerte gebildeten Rangreihen der Klassen (vgl. Tabelle 39).

Tabelle 39: Mittelwerte ($\bar{X}$) und Rangplätze (R) der 5 Klassen in der Versuchsform VI, K und in dem Verfahren von Wallach und Kogan

		Klasse				
		A	B	C	D	E
Versuchsform VI, K	$\bar{X}$	53.07	45.93	60.57	65.87	67.28
	R	4	5	3	2	1
Kreativitätstest von Wallach und Kogan	$\bar{X}$	167.09	137.20	256.82	185.53	160.72
	R	4	2	1	3	5

Außer der Klasse A behielt keine der Klassen ihren Rangplatz von der ersten Testung bei. Daß es zu so großen Rangplatzverschiebungen kommen würde wie bei den Klassen B und vor allem E, war nicht erwartet worden und ist ohne Informationen über Veränderungen in der Motivationslage der Schüler oder in den Durchführungsbedingungen der beiden Tests

nicht erklärbar.[1] Falls weitere Untersuchungen dieser Art auch bei stärkerer Kontrolle der Testsituation zu so stark differierenden Ergebnissen kommen sollten, müßte die Frage der zeitlichen Stabilität der Verfahren getrennt von derjenigen ihrer Beeinflußbarkeit durch situationale Bedingungen untersucht werden.

Eine zusammenfassende Beurteilung dieser Untersuchung der konvergenten und diskriminanten Validität der Versuchsform VI, K kommt zu einem positiven Ergebnis. Vor allem die klassenspezifisch berechneten Korrelationskoeffizienten weisen auf eine angesichts der Verschiedenheit der beiden Verfahren enge Beziehung zwischen den beiden Kreativitätstests hin, während ihre Zusammenhänge mit Intelligenztestwerten signifikant niedriger sind.
Die Ergebnisse dieser Untersuchung enthalten eventuell Hinweise auf eine relativ starke Abhängigkeit der Testwerte von der jeweiligen Untersuchungssituation. Weitere Untersuchungen müßten klären, ob diese Hypothese zutrifft, ob sie für beide Verfahren oder nur für eines von beiden Gültigkeit besitzt, und wenn ja, für welches.

3.3325 Ergebnisse der Untersuchung mit dem Persönlichkeitsfragebogen CPQ

In für Jungen und Mädchen getrennt durchgeführten multiplen Regressionsanalysen[2] wurde versucht, die Ergebnisse der Versuchsform VI, K mit Hilfe des Persönlichkeitsfragebogens CPQ vorherzusagen. Da der CPQ nicht für deutsche Kinder adaptiert ist, sollte durch Wiederholung dieser Prozedur mit dem Kreativitätstest von Wallach und Kogan als abhängige Variable die Interpretation eines den Hypothesen widersprechenden Ergebnisses abgesichert werden: wenn die Gesamtwerte beider Verfahren nicht durch die hypothetischen "Kreativitätsfaktoren" des CPQ vorhergesagt werden, sollte dies nicht als Nachweis der mangelnden Validität der beiden Tests, sondern als Hinweis auf die mangelnde Gültigkeit des CPQ für deutsche Schüler gewertet werden.[3]

---

1) Daß allein die Aussicht auf eine Turnstunde, eine Filmvorführung, einen Unterrichtsgang o.a. die auffallende Rangplatzverschlechterung der Klasse E bewirkt haben könnte, ist angesichts der Ergebnisse von Elkind, Deblinger und Adler (1970) nicht unwahrscheinlich (vgl. Tabelle 4, S. 119).

2) Programm REGT des Deutschen Rechenzentrums; Autor: Dr. F. Gebhardt, DRZ Darmstadt

3) Die Hypothesen über Richtung und Stärke der Beziehung von 7 CPQ-Faktoren zu den Kreativitätstestwerten sind auf den Seiten 208-211 dargestellt.

Tabelle 46: Multiple Korrelationen (R) und Populationsschätzwerte der multiplen Korrelationen ($\bar{R}$)[1] der CPQ-Faktoren mit den Produktivitätswerten zweier Kreativitätstests

	Ergebnisse der Jungen (N = 67)			Ergebnisse der Mädchen (N = 62)		
	R	$\bar{R}$	Prädiktoren	R	$\bar{R}$	Prädiktoren
Versuchsform VI,K	0.569	0.55o	E, C, F*	0.4o8	0.37o	B, C, - E*
	0.621	0.58o	E, C, - G, F, - Q_4, - J	0.577	0.495	B, - Q_3, C, F, G, - Q_4 - E, J,
	0.631	0.5o1	alle 14 CPQ-Faktoren	0.588	0.411	alle 14 CPQ-Faktoren
	Ergebnisse der Jungen (N = 6o)			Ergebnisse der Mädchen (N = 54)		
	R	$\bar{R}$	Prädiktoren	R	$\bar{R}$	Prädiktoren
Kreativitätstest von Wallach und Kogan	0.552	0.529	- G, C, B	0.398	0.354	C, - F, B
	0.634	0.567	- G, C, A, B, F, Q_3, - D, - H	0.496	0.41o	C, B, - F, G, - Q_4, N
	0.64o	0.493	alle 14 CPQ-Faktoren	0.516	0.17o	alle 14 CPQ-Faktoren

*) Die Reihenfolge der CPQ-Faktoren ist durch die Höhe der Standardregressionskoeffizienten β determiniert. Negative Vorzeichen bedeuten negative Standardregressionskoeffizienten. Den Hypothesen entsprechende Ergebnisse wurden unterstrichen.

1) S. dazu S. 176

Für die Stichprobe der Jungen werden die Ergebnisse der beiden Kreativitätstests durch die CPQ-Faktoren gut vorhergesagt. Für die Versuchsform VI, K wurden die Untersuchungshypothesen insofern bestätigt, als die Faktoren E und C die beiden wichtigsten Prädiktoren sind. Der drittbeste Prädiktor ist Faktor F und nicht - wie erwartet - B, so daß sich für die Jungen die Versuchsform VI, K als ein Verfahren darstellt, das von der allgemeinen Intelligenz, wie sie durch den Faktor B des CPQ erfaßt wird, überraschend unabhängig ist.

Die Prädiktorenkombination, die zu dem höchsten Populationsschätzwert der multiplen Korrelation führt, umfaßt außerdem den Faktor O mit dem erwarteten Vorzeichen. Zu den Faktoren G und Q_4, die ebenfalls zur Vorhersage der VI, K-Ergebnisse beitragen, wurden keine Hypothesen aufgestellt. Daß G mit negativer Gewichtung in die Regressionsgleichung eingeht, widerspricht der Spezifikationsgleichung von Cattell und Butcher und einem Teil der Untersuchungsergebnisse (vgl. Tabelle III in Anhang B 2). An der Bedeutung des Faktors G für die Kreativität von Schulkindern wurden allerdings bereits Zweifel geäußert.[1)]

Die negative Gewichtung von G erfährt eine Bestätigung durch die multiplen Regressionsanalysen zur Vorhersage der Ergebnisse des Tests von Wallach und Kogan, in denen G - ebenfalls mit negativer Gewichtung - der wichtigste Prädiktor ist (s.u.)

Die Prädiktorenkombination zur Vorhersage der Ergebnisse des Verfahrens von Wallach und Kogan stimmen etwas weniger gut mit den Hypothesen über die Beziehungen der CPQ-Faktoren zu den Kreativitätstestwerten überein.

Völlig unerwartet ist gerade bei diesem aus 39 Items bestehenden Verfahren die negative und sehr starke Gewichtung des Faktors G. Demnach erhalten nicht die gewissenhaften, ausdauernden und leistungsmotivierten Jungen hohe Werte in dem Test von Wallach und Kogan, sondern diejenigen mit geringer Überich-Stärke.[2)] Die in dieser Arbeit und von Guilford vertretene Hypothese[3)], daß ein hohes Maß an Anstrengungsbereitschaft Voraussetzung für ein gutes Testergebnis ist, wird durch dieses Resultat widerlegt oder jedenfalls in Frage gestellt.

---

1) S. dazu S. 207

2) Vgl. Tabelle 2o, S. 2o9

3) S. dazu S. 142 und S. 86

Abgesehen von der unerwartet starken Gewichtung des Faktors G erhält in der Regressionsanalyse mit 8 Prädiktoren der Faktor H im Gegensatz zur Hypothese eine negative Gewichtung. Zu den Faktoren A, D und Q_3, die ebenfalls zu diesen 8 Prädiktorvariablen gehören, wurde keine Hypothese aufgestellt; daß Faktor A als drittbester Prädiktor mit positivem Vorzeichen in die Regressionsgleichung eingeht, ist etwas überraschend. Auch die gegensätzliche Gewichtung von G und Q_3 ist schwer zu interpretieren. Vor allem aber ist es erstaunlich, daß der für kreatives Verhalten sehr bedeutsame Faktor E nicht zu den 8 wichtigsten Prädiktoren gehört.

Die Kreativitätstestergebnisse der Mädchen werden - wie erwartet - durch den CPQ weniger gut vorhergesagt. Die Faktoren B und C sind die wichtigsten Prädiktoren.
Der Faktor E hat in der deutschen Stichprobe bei den Mädchen eine negative Gewichtung und bestätigt somit die theoretische Erwartung über die Auswirkung der geschlechtsrollenspezifischen Erziehung (s. Tabelle 2o, S. 2o8) und einige amerikanische Untersuchungsergebnisse (vgl. Tabelle III, Anhang B 2). Wie bei den Regressionsanalysen zur Vorhersage der Ergebnisse des Tests von Wallach und Kogan für die Jungen haben die Faktoren G und Q_3 Gewichtungen mit entgegengesetzten Vorzeichen, aber bei den Mädchen ist eine hohe Überich-Stärke zusammen mit niedrigen Q_3-Werten ein Hinweis auf bessere Resultate der Versuchsform VI, K.
Auch die Resultate des Kreativitätstests von Wallach und Kogan werden durch die Faktoren B und C am besten vorhergesagt; außerdem spielt Faktor F, aber mit negativem Vorzeichen, eine wichtige Rolle. Da für die Mädchen keine Hypothesen aufgestellt wurden, werden diese Resultate nicht weiter diskutiert.

Aufgrund der Ergebnisse der Jungen kann die Frage, ob die mit der Versuchsform VI, K ermittelten Werte durch die Persönlichkeitsfaktoren des CPQ so gut vorhergesagt werden können, daß daraus ein Validitätsbeleg abgeleitet werden kann, bejaht werden. Die Höhe der mit 3 Prädiktoren erreichten multiplen Korrelation kann als Hinweis auf die Konstruktvalidität dieses Verfahrens zur Erfassung des divergenten Denkens betrachtet werden.
Die Vorhersage der Ergebnisse des Intelligenztests P-S-B durch die CPQ-Faktoren wird nun noch mit derjenigen der VI,K-Resultate verglichen.

Tabelle 41 zeigt die Ergebnisse der multiplen Regressionsanalysen. Die multiplen Korrelationen zwischen den CPQ-Faktoren und den Gesamtwerten der Versuchsform VI, K werden für die jeweils gleiche Prädiktorenzahl in Klammern mitgeteilt.

Tabelle 41: Multiple Korrelationen (R) und Populationsschätzwerte der multiplen Korrelationen ($\bar{R}$)[1] der CPQ-Faktoren mit den P-S-B-Gesamtwerten

	Ergebnisse der Jungen (N = 63)			Ergebnisse der Mädchen (N = 5o)		
	R	$\bar{R}$	Prädiktoren	R	$\bar{R}$	Prädiktoren
P-S-B-Gesamtwert	o.5o1	o.476	B, C, - O*	o.454	o.415	C, - A, B
	(o.569)	(o.55o)	(E, C, F)	(o.4o8)	(o.372)	(B, C, - E)
	o.551	o.493	C, B, A, - O, - H, D	o.51o	o.442	C, B, - A, G, - H
	(o.621)	(o.58o)	(E, C, -G, F, - Q_4, - O)	(o.491)	(o.434)	(B, C, F, - E, - Q_3)
	o.594	o.426	**alle 14 CPQ-Faktoren**	o.554	o.238	**alle 14 CPQ-Faktoren**
	(o.631)	(o.5o1)	alle 14 CPQ-Faktoren	(o.58ß)	(o.411)	alle 14 CPQ-Faktoren

*) Die Reihenfolge der CPQ-Faktoren ist durch die Höhe der Standardregressionskoeffizienten β determiniert. Negative Vorzeichen bedeuten negative Standardregressionskoeffizienten. In Klammern werden die Werte für R und $\bar{R}$ und die Prädiktoren der VI,K-Ergebnisse mitgeteilt.

1) S. dazu S. 176

Die P-S-B-Ergebnisse der Jungen werden durch die CPQ-Faktoren weniger gut vorhergesagt als die Gesamtwerte der Versuchsform VI, K. Die Signifikanz des Unterschieds zwischen den multiplen Korrelationskoeffizienten kann allerdings nicht geprüft werden, da für dieses Problem kein Signifikanztest vorzuliegen scheint.[1] Dieses Resultat gewinnt dadurch besonderes Gewicht, daß die B-Skala als Maß der allgemeinen Intelligenz einer der Prädiktoren des Intelligenztestwertes ist. Erwartungsgemäß sagt sie das P-S-B-Ergebnis am besten vorher.[2] Daß die Vorhersage trotzdem schlechter ist als diejenige des Gesamtwerts der Versuchsform VI, K, ist ein deutlicher Hinweis auf die Abhängigkeit des letzteren von nicht-intellektuellen Persönlichkeitsvariablen.

Bei den Mädchen, für die keine Hypothesen formuliert wurden, sagen die CPQ-Faktoren das Intelligenztestergebnis bei den Regressionsanalysen mit geringeren Prädiktorenzahlen besser vorher als die Kreativitätstestwerte. Faktor C ist bei den Mädchen für die Vorhersage des P-S-B-Ergebnisses wichtiger als Faktor B.

Die Ergebnisse der Untersuchung mit dem CPQ an den Jungen der Teilstichprobe können als Beleg für die Konstruktvalidität der Versuchsform VI, K gewertet werden. Für die Faktoren B, C und E waren die engsten Zusammenhänge mit den Gesamtwerten von VI, K erwartet worden. Für C und E wurde diese Erwartung bestätigt; drittbester Prädiktor war nicht B, sondern F.

Die Hypothese, daß die Ergebnisse des P-S-B durch die CPQ-Faktoren weniger gut vorhergesagt werden als die Gesamtwerte der Versuchsform VI, K, wurde ebenso bestätigt wie die Erwartung, daß die Zusammenhänge zwischen den CPQ-Faktoren und den Kreativitätstestergebnissen bei den Mädchen weniger eng sind als bei den Jungen.

---

1) Da hier 2 verschiedene Kriterien durch die gleichen Prädiktoren vorhergesagt werden, wäre die kanonische Korrelation das angemessene Verfahren. Nachdem jedoch der Vergleich der beiden multiplen Korrelationen nur im Fall zweifelhafter Ergebnisse der multiplen Regressionsanalyse zur Vorhersage der VI, K-Werte als Entscheidungshilfe für die Beurteilung der Untersuchungsresultate herangezogen werden sollte, erschien die Berechnung der kanonischen Korrelation nicht unbedingt erforderlich.

2) Sie ist der Prädiktor der letzten eingeschränkten Hypothese der Regressionsanalyse.

4 Zusammenfassung der Arbeit und Wertung der Ergebnisse

Die vorliegende Arbeit hatte die Entwicklung oder Adaptierung eines Verfahrens zur Untersuchung kognitiver Voraussetzungen für kreatives Verhalten für 9- bis 12 jährige Schüler zum Ziel.

Aufgrund einer eingehenden Analyse der Literatur zu verschiedenen Kreativitätskonzeptionen und zu den für jüngere Schüler geeigneten Verfahren erschien es nicht angezeigt, einen der bereits vorliegenden und erprobten amerikanischen Kreativitätstests zu adaptieren. Die Kreativitätskonzeptionen und die mit den Kreativitätstests gemachten Erfahrungen wurden jedoch für die theoretische Fundierung, Entwicklung und Validierung des eigenen Verfahrens nutzbar gemacht.

Guilfords Konzept des divergenten Denkens und die von Wallach und Kogan sowie von Krause weiterentwickelte Kreativitätstheorie Mednicks sind die Ausgangspunkte der theoretischen Konzeption des Verfahrens.

Die Berücksichtigung der Theorie Guilfords führte zur Entwicklung von Aufgaben mit freier Beantwortungsform und zu der Zielsetzung, daß das Verfahren divergentes Denken in mehr als einem "Inhaltsbereich" im Sinne Guilfords erfassen soll.

Die Überlegungen der assoziationstheoretisch orientierten Kreativitätsforscher wurden insofern berücksichtigt, als die Aufgaben den Probanden nur minimale Beschränkungen auferlegen, so daß die meisten Kinder mehrere Lösungen finden können. Im Interesse einer entspannten Untersuchungssituation wurde die Zeit für die Aufgabenbearbeitung großzügig bemessen und die Einhaltung der Zeitgrenzen möglichst unmerklich realisiert.

Die assoziationstheoretische Kreativitätskonzeption führte außerdem zu der Hypothese einer relativ engen Beziehung zwischen den Ergebnissen des zu entwickelnden Verfahrens und nicht-intellektuellen Persönlichkeitsvariablen.

Im Gegensatz zu Wallach und Kogan, aber in Übereinstimmung mit Krause wurde ein mäßiger positiver Zusammenhang mit Intelligenztestergebnissen erwartet.

68 Einzelaufgaben, die entweder aus amerikanischen Kreativitätstests - meist in modifizierter Form - übernommen oder in Anlehnung an bekannte Kreativitätstest-Items konstruiert worden waren, wurden zunächst in face-to-face-Testungen erprobt. 58 dieser Aufgaben wurden auf die ersten beiden Versuchsformen verteilt und an 312 bzw. 264 Schülern vierter Klassen erprobt.
Die Antworten der Kinder wurden nach der Anzahl akzeptabler Lösungen ("Produktivität") und nach "Originalität" im Sinn von statistischer Seltenheit ausgewertet.
Die wichtigsten Beurteilungskriterien für die Testaufgaben waren in diesem frühen Stadium der Testentwicklung ihre Trennschärfe und eine weitgehende Unabhängigkeit der Ergebnisse vom Geschlecht und von der Sozialschicht der Probanden.
Diejenigen 11 Aufgaben, die nach diesen Kriterien am besten abgeschnitten hatten, wurden zur Versuchsform IV, 1 zusammengestellt.
Diese Versuchsform wurde an 526 Schülern aus 5. Gesamtschul-Klassen und an 130 Hauptschülern der 5. Klassenstufe erprobt.
Die Untersuchung in den Gesamtschulklassen diente vor allem der Überprüfung der Trennschärfe und internen Konsistenz des Verfahrens und dem Nachweis, daß die Aufgaben der Versuchsform IV, 1 nicht nur aufgrund ihrer Beziehung zur Testintelligenz zusammenhängen, sondern ein von dieser abgrenzbares Persönlichkeitsmerkmal messen.
Die Ergebnisse waren insgesamt befriedigend und führten lediglich zu kleineren Modifikationen des Verfahrens.
Die Untersuchung in den 4 Hauptschulklassen diente der Überprüfung der Übereinstimmungsvalidität und erbrachte weniger gute Resultate. Da es sich um 4 Klassen aus derselben Schule handelte, war allerdings nicht auszuschließen, daß die unbefriedigenden Ergebnisse auf Besonderheiten der Stichprobe zurückzuführen sind.
Mit der Revision der Versuchsform IV, 1 wird die Testentwicklung im wesentlichen als abgeschlossen betrachtet.

Die Untersuchungen mit der letzten Versuchsform - VI, K - dienten in erster Linie der Testvalidierung. Neben Überprüfungen bestimmter Voraussetzungen für die Validität des Verfahrens - die Unabhängigkeit des Verfahrens vom Geschlecht und der Sozialschicht der Probanden sowie die diagnostische Valenz der Originalitätswerte - wurden 2 Untersuchungen zur Übereinstimmungsvalidität und 4 Untersuchungen zur Konstruktvalidität durchgeführt.

Sie führten zu folgenden Ergebnissen:

1. Die Mädchen erreichen in den meisten Aufgaben und in den Produktivitäts- und Originalitäts-Gesamtwerten des Verfahrens bessere Ergebnisse als die Jungen, so daß eine geschlechtsspezifische Normierung angezeigt erscheint. Die Frage der differentiellen Validität muß weiter untersucht werden.

2. Der schichtspezifische Bias des Verfahrens ist sehr gering; seine Auswirkungen können vernachlässigt werden.

3. Aufgrund der Korrelation von .92 zwischen den Produktivitäts- und Originalitäts-Gesamtwerten erscheint die Auswertung der Testergebnisse nach ihrer statistischen Seltenheit überflüssig. Das Verfahren wird daher nur noch nach der Zahl der akzeptablen Antworten ausgewertet.

4. In einer Untersuchung zur Übereinstimmungsvalidität wurden die Klassenlehrer von 5 Gesamtschul-Klassen über Verhaltensweisen der Schüler befragt, die Rückschlüsse auf die Motivation für kreatives Verhalten ermöglichen.
   Die Korrelation der Fragebogenergebnisse mit dem Gesamtwert der Versuchsform VI, K war sehr niedrig; für eine ganze Reihe schulisch relevanter Variablen wurden erheblich engere Zusammenhänge gefunden.
   Die Zusammenhänge zwischen den Ergebnissen des Lehrerfragebogens und den anderen Variablen sind von Klasse zu Klasse sehr verschieden, was wahrscheinlich bedeutet, daß jeder der 5 Lehrer bei der Beantwortung der Fragen andere Beurteilungskriterien verwendete.
   Aufgrund der Ergebnisse dieser Untersuchung kann somit die Übereinstimmungsvalidität der Versuchsform VI, K nicht beurteilt werden.

5. In einer zweiten Untersuchung zur Übereinstimmungsvalidität des Verfahrens wurden Schülerratings als Außenkriterium verwendet. Auch hier ergab sich eine sehr niedrige Korrelation mit der Versuchsform VI, K und einige mäßige bis mittlere Zusammenhänge mit einigen unterrichtsrelevanten Schülervariablen und dem Ergebnis des Tests von Wallach und Kogan.
   Klassenweise Berechnungen der Korrelationen erbrachten wieder für jede der 5 Klassen ein anderes Bild.
   Somit lassen auch die Resultate der Schülerbefragung die Untersuchungsfrage unbeantwortet.

Die Ergebnisse der beiden Untersuchungen zur Übereinstimmungsvalidität lassen es äußerst fraglich erscheinen, ob klassenweise erhobene Schätzurteile von Lehrern und/oder Schülern geeignete Außenkriterien für einen Test zur Untersuchung der Fähigkeiten des divergenten Denkens sind.

6. Drei Untersuchungen zur Konstruktvalidität der Versuchsform VI, K überprüften die interne Konsistenz und die konvergente und diskriminante Gültigkeit des Verfahrens. Nach den Ergebnissen erfassen die Testaufgaben ein faktoriell komplexes Persönlichkeitsmerkmal, dessen Dimensionen von den durch Intelligenztest-, Schulleistungs-, Angstfragebogen- und Sozialschicht-Variablen repräsentierten Faktoren deutlich abgegrenzt sind.

   Die interne Konsistenz der Aufgaben der Versuchsform VI, K ist ausreichend hoch, ihre Korrelationen mit Intelligenztestwerten sind deutlich niedriger, so daß die diskriminante Validität des Verfahrens als gut bezeichnet werden kann.
   Eine Überprüfung der konvergenten und diskriminanten Validität des Tests mit Hilfe einer vereinfachten "multitrait-multimethod matrix" erbrachte bei klassenweise berechneten Korrelationen zwischen der Versuchsform VI,K,dem Kreativitätstest von Wallach und Kogan und 2 Intelligenztests sehr gute Ergebnisse: Die beiden Kreativitätstests interkorrelieren etwa ebenso hoch wie die beiden Intelligenztests; diese Koeffizienten sind signifikant bzw. hochsignifikant höher als die 4 Korrelationen zwischen Kreativitäts- und Intelligenztests.
7. Im Rahmen der Konstruktvalidierung wurden auch die Zusammenhänge zwischen den Ergebnissen der Versuchsform VI, K und mehreren für kreatives Verhalten bedeutsamen Persönlichkeitsvariablen untersucht.
   Die Hypothesen über Vorzeichen und relative Gewichtung der entsprechenden Faktoren des "Children's Personality Questionnaire" von Porter und Cattell wurden durch die Ergebnisse einer multiplen Regressionsanalyse im wesentlichen bestätigt.
   Erwartungsgemäß wurden die VI, K-Ergebnisse der Jungen besser vorhergesagt als die der Mädchen.

Eine zusammenfassende Beurteilung aller Untersuchungsergebnisse kommt zu dem Schluß, daß die Validität des neu entwickelten Verfahrens zur Messung des divergenten Denkens zwar nicht ausreichend belegt werden konnte, daß die bereits vorliegenden Resultate jedoch für eine Fortführung der Validierungsarbeiten sprechen.

LITERATURVERZEICHNIS

Amelang, M. und Vagt, G. Warum sind Schulnoten von Mädchen durch Leistungstests besser vorherzusagen als diejenigen von Jungen? Z. Entw. u. Päd. Psychol. 1970, 2, S. 210 - 220

Arasteh, J. D. Creativity and related processes in the young child: a review of the literature. J. gen. Psychol. 112 (1968), S. 77 - 108

Aurin, K. Ermittlung und Erschließung von Begabungen im ländlichen Raum; In: Bildung in neuer Sicht, Schriftenreihe des Kultusministeriums Baden-Württemberg zur Bildungsforschung. 1966, Reihe A, Nr. 2

Bachtold, L. M. The creative personality and the ideal pupil revisited. J. creat. Behav. 8 (1974) 1, S. 47 - 54

Bader, L. J. The effects of task-taking atmosphere, level of creativity, and field-independance on creative production. Diss. Abstr. 1970, 31, 5-B, S. 2974

Barabasz, A. F. Test-retest reliability of the Torrance Tests of Creative Thinking, and the relationship between intelligence and figural creativity. In: Child Study Center Bulletin, State University Coll., New York, Buffalo 5 (1969), 4 - 5, S. 73 - 74

Barron, F. Some personality correlates of independence of judgment. J. Pers. 21 (1953), S. 287 - 297

Barron, F. Creative vision and expression in writing and painting. In: Conference on the Creative Person. Berkeley: University of California, Institute of Personality Assessment and Research 1961, Ch. 2

Barron, F. Creative person and creative process. New York: Holt, Rinehart and Winston 1969

Bartlett, M. M. and Davis, G. A. Do the Wallach and Kogan tests predict real creative behavior? Percept. mot. Skills 39 (1974), S. 730

Bauer, A. Zum sozio-ökonomischen und sozio-kulturellen Bildungsmilieu. In: Mitteilungen und Nachrichten des Deutschen Instituts für Internationale Pädagogische Forschung, Sonderheft 1971

Beck, O. und Arbeitsgruppe "Pädagogische Diagnostik" an der Erziehungswissenschaftlichen Hochschule Rheinland-Pfalz. Kriterien zur Aufsatzbeurteilung. Mainz: V. Hase und Koehler 1974

Begemann, E. Kreativität und Schule. Die Deutsche Schule 64 (1972) 6, S. 362 - 376

Bennett, M. W. Factors influencing performance on group and individual tests of intelligence: I. Rate of work. Genetic psychol. Monogr. 23 (1941), S. 237 - 318

Bennett, S. N. Divergent thinking abilities - a validation study. Brit. J. educ. Psychol. 43 (1973) 1, S. 1 - 7

Bhavnani, R., and Hutt, C. Divergent thinking in boys and girls. J. Child Psychol. Psychiatr. 13 (1972), S. 121 - 127

Bigelow, G. S. Field dependence - field independence in 5- to 10-year-old children. J. educ. Res. 64 (1971) 9, S. 397 - 400

Bloomberg, M. Creativity as related to field independence and mobility. J. gen. Psychol. 118 (1971), S. 3 - 12

Boersma, F. J., and O'Bryan, K. An investigation of the relationship between creativity and intelligence under two conditions of testing. J. Pers. 36 (1968), S. 341 - 348

Brickenkamp, R. Kreativität und Kreativitätstraining in der Schule. In: Nickel, H. und Langhorst, E. (Hrsg.) Brennpunkte der pädagogischen Psychologie. Bern: Hans Huber 1973

Brühl, G. Kreativität und Schule. Lebendige Schule 3 (1970), S. 81 - 87

Buggle, F. und Baumgärtel, F. Hamburger Neurotizismus- und Extraversionsskala für Kinder und Jugendliche HANES, KJ. Göttingen: Hogrefe 1972

Butcher, H.J. Intelligence and creativity. In: Kline, P. (Ed.) New approaches in psychological measurement. New York: Wiley 1973, S. 43 - 64

Carroll, J. B. Guilford, J. P. The nature of human intelligence. Amer. educ. Res. J. 5 (1968) 2, S. 249 - 256

Carroll, J. B. Stalking the wayward factors. Contemp. Psychol. 17 (1972) 6, S. 321 - 324

Cattell, R. B. The personality and motivation of the researcher from measurements of contemporaries and from biography. University of Utah Research Conference for the Identification of Creative Scientific Talent, 1959, S. 77 - 93

Cattell, R. B.,and Butcher, H. J. The prediction of achievement and creativity, New York: Bobbs-Merrill Company, Inc. 1968

Cattell, R. B.,and Drevdahl, J. E. A comparison of the personality profile (16PF) of eminent researchers with that of eminent teachers and administrators and of the general population. Brit. J. Psychol. 46 (1955), S. 248 - 261

Christensen, P. R., Guilford, J. P. and Wilson, R. C. Relations of creative responses to working time and instructions. J. exp. Psychol. 53 (1957) 2, S. 82 - 88

Christie, T. Environmental factors in creativity. J. creat. Behav. 4 (1970) 1, S. 13 - 31

Cicirelli, V. G. Form of the relationship between creativity, IQ, and academic achievement. J. educ. Psychol. 56 (1965) 6, S. 303 - 308

Cicirelli, V. G. Sibling constellation, creativity, IQ, and academic achievement. Child Developm. 38 (1967), S. 481 - 490

Cline, V. B., Richards, J. M., Jr., and Abe, C. The validity of a battery of creativity tests in a high school sample. Educ. psychol. Measmt. 22 (1962), S. 781 - 784

Cline, V. B., Richards, J. M., Jr., and Needham, W. E. Creativity tests and achievement in high school science. J. appl. Psychol. 47 (1963) 3, S. 184 - 189

Crockenberg, S. B. Creativity tests: a boon or boondoggle for education? Rev. educ. Res. 42 (1972), S. 27 - 45

Cronbach, L. J. Intelligence? Creativity? A parsimonious reinterpretation of the Wallach-Kogan data. Amer. educ. Res. J. 5 (1968) 4, S. 491 - 511

Cronbach, L. J. Test validation. In: Thorndike, R. L. (Ed.) Educational Measurement. Washington, D. C.: American Council on Education 1971, S. 443 - 507

Cronbach, L. J., and Meehl, P. E. Construct validity in psychological tests. Psychol. Bull. 52 (1955) 4, S. 281 - 302

Cropley, A. J. Creativity and intelligence. Brit. J. educ. Psychol. 36 (1966) 3, S. 259 - 266

Cropley, A. J. A note on the Wallach-Kogan tests of creativity. Brit. J. educ. Psychol. 38 (1968), S. 197 - 201

Cropley, A. J. A five-year longitudinal study of the validity of creativity tests. Developm. Psychol. 6 (1972) 1, S. 119 - 124

Cropley, A.J., and Maslany, G.W. Reliability and factorial validity of the Wallach-Kogan creativity tests. Brit. J. Psychol. 6o (1969) 3, S. 395 - 398

Crutchfield, R.S. The creative process.In: Conference on the Creative Person. Berkeley University of California, Institute of Personality Assessment and Research, 1961, Ch. 6

Crutchfield, R.S. Detrimental effects of conformity pressure on creative thinking. Psychol. Beitr. 6 (1962), S. 463-471

Crutchfield, R.S. Conformity and creative thinking. In: Gruber, H.E., Terrell, G., and Wertheimer, M. (Eds.) Contemporary approaches to creative thinking. New York: Atherton Press 1963, S. 12o - 14o

Dacey, J., and Ripple, R. Relationships of some adolescent characteristics and verbal creativity. Psychology in the Schools 6 (1969) 3, S. 321 - 324

Dacey, J.S., and Madaus, G.F. An analysis of two hypotheses concerning the relationship between creativity and intelligence. J.educ. Res. 64 (1971) 5, S. 213-216

Datta, L.E., and Parloff, M.B. On the relevance of autonomy: parent-child relationships and early scientific creativity. Proceedings, 75 th Annual Convention, APA, 1967 (reprint)

Davidson, W.M., and Carrol, J.B. Speed and level components in time-limit scores: a factor analysis. Educ. psychol. Measmt. 5(1945) , S. 411 - 427

Dentler, R.A., and Mackler, B. Originality: some social and personal determinants. Behavior. Science 9 (1964), S. 1 - 7

Dewing, K. The reliability and validity of selected tests of creative thinking in a sample of seventhgrade West Australian children. Brit. J. educ. Psychol. 4o (197oa) 1, S. 35-42

Dewing, K. Some correlates of creativity test performance in seventh grade children. In: Australian J. Psychol. 22 (197ob) 3, S. 269-276

Dewing, K. Family influences on creativity: a review and discussion. J. spec. Educ. 4 (197o) 4, S. 399-4o4

Domino, G. Maternal personality correlates of sons' creativity. J. consult. clin. Psychol. 33 (1969) 2, S. 18o - 183

Drevdahl, J.E. Some developmental and environmental factors in creativity. In: Taylor, C.W. (Ed.) Widening horizons in creativity. New York: Wiley 1964, S. 17o-186

Drevdahl, J.E., and Cattell, R.B. Personality and creativity in artists and writers. J. clin. Psychol. 14 (1958), S. 1o7-111

Edwards, A.L. Versuchsplanung in der psychologischen Forschung. Weinheim: Beltz 1971

Edwards, M.P., and Tyler, L.E. Intelligence, creativity, and achievement in a nonselective public junior high school. J. educ. Psychol. 56 (1965), S. 96-99

Elkind, D., Deblinger, J., and Adler, D. Motivation and creativity: the context effect. Amer.educ. Res. J. 7 (197o) 3, S. 351-357

English, H.B., and English, A.C. A comprehensive dictionary of psychological and psychoanalytical terms. London: Longmans 1958

Eysenck, H.J. Intelligence assessment: a theoretical and experimental approach. Brit. J. educ. Psychol. 37 (1967), S. 81 - 98

Fee, F. An alternative to Ward's factor analysis of Wallach and Kogan's " Creativity " correlations. Brit. J. educ. Psychol. 38 (1968), S. 319 - 321

Fiedler, P.A., und Windheuser, H.J. Modifikation kreativen Verhaltens durch Lernen am Modell. Z. Entw. u. Päd. Psychol. 6 (1974) 4, S. 262-28o

Fippinger, F. Allgemeiner Schulleistungstest für 4. Klassen (AST 4). Weinheim:Beltz 1967

Flechsig, K.H. Erziehen zur Kreativität. Neue Sammlung 6 (1966), S. 129 - 141

Fleming, E.S., and Weintraub, S. Attitudinal rigidity as a measure of creativity in gifted children. J. educ. Psychol. 53 (1962) 2, S. 81-85

Flescher, J. Anxiety and achievement of intellectually and creatively gifted children. J. Psychol. 56 (1963), S. 251-268. Deutsch in: Mühle, G. und Schell, C. (Hrsg.) Kreativität und Schule. München:Piper 197o

Freeman, F.S. Theory and practice of psychological testing. New York 1966

Freeman, J., Butcher, H.J., and Christie, T. Creativity - a selective review of research. London 1968

Fürntratt, E. Zur Bestimmung der Anzahl interpretierbarer gemeinsamer Faktoren in Faktorenanalysen psychologischer Daten. Diagnostica 15 (1969) 1, S. 62 - 75

Gärtner-Harnach, V. Fragebogen für Schüler F S. Hrsg.: K. Ingenkamp. Aus der Reihe " Deutsche Schultests ". Weinheim: Beltz 1973

Gastel, A. Über Sprachmängel bei Schulanfängern. Unveröffentlichte Dissertation der Med. Fakultät. München 1964

Getzels, J.W., and Jackson, P.W. The meaning of "giftedness". An examination fo an expanding concept. Phi Delta Kappan 4o (1958), S. 75-77. Deutsch in: Mühle, G. und Schell, C. (Hrsg.) Kreativität und Schule. München:Piper 197o

Getzels, J.W., and Jackson, P.W. Creativity and intelligence. New York : Wiley 1962

Goldmann, R.J. The Minnesota Tests of Creative Thinking. Educ. Res. 7 (1964), S. 3 - 14

Gough, H.G. Techniques for identifying the creative research scientist. In : Conference on the Creative Person. Berkeley : University of California , Institute of Personality Assessment and Research 1961, Ch. 3

Grimm, H. Stukturanalytische Untersuchung der Kindersprache. Bern : Hans Huber 1973

Gross, H. (Hrsg.) Zukunft aus Kreativität. Düsseldorf, Wien : Econ 1971

Grote, J., Hajek, G. et al. Untersuchungen zur Kreativität. Z. f. Päd. 15 (1969) 2, S. 135 - 171

Guilford, J.P. Kreativität (195o). In: Mühle, G.und Schell, C.(Hrsg.) Kreativität und Schule. München: Piper 197o, S. 13 - 36

Guilford, J.P. Zero correlations among tests of intellectual abilities. Psychol. Bull. 61 (1964) 6, S. 4o1-4o4

Guilford, J.P. Intellectual factors in productive thinking. In: Mooney, R.L., and Razik, T.A. (Eds.) Explorations in creativity. New York:Harper and Row 1967 a, S. 95 - 1o6

Guilford, J.P. The nature of human intelligence. New York: Mc Graw-Hill 1967 b

Guilford, J.P. Measurement of creativity. In: Mooney, R.L., and Razik, T.A. (Eds.) Explorations in creativity. New York : Harper and Row 1967 c, S. 281-287

Guilford, J.P. Persönlichkeit. Weinheim:Beltz 197o[4]

Guilford, J.P. Creativity: retrospect and prospect. J. creat. Behav. 4 (197o) 3, S. 149 - 168

Guilford, J.P. Some misconceptions regarding measurement of creative talents. J. creat. Behav. 5 (1971) 2, S. 77-87

Guilford, J.P.,and Christensen, P.R. A factor-analytic study of verbal fluency. Rep. psychol. Lab., No. 17. Los Angeles: University of Southern California 1956

Guilford, J.P., and Christensen, P.R. The one-way relation between creative potential and IQ. J. creat. Behav. 7 (1973) 4, S. 247 - 252

Guilford, J.P., and Hoepfner, R. Sixteen divergent-production abilities at the ninth-grade level. Multivariate Behavioral Research 1966, 1, S. 43-66

Guilford, J.P., and Hoepfner, R. The analysis of intelligence. New York : Mc Graw-Hill 1971

Haddon, F.A., and Lytton, H. Teaching approach and the development of divergent thinking abilities in primary schools. Brit. J. educ. Psychol. 38 (1968), S. 171-18o

Haddon, F.A., and Lytton, H. Primary education and divergent thinking abilities - four years on. Brit. J. educ. Psychol. 41 (1971), S. 136 - 147

Hargreaves, D.J.,and Bolton, N. Selecting creativity tests for use in research. Brit. J. Psychol. 63 (1972) 3, S. 451-462

Harvey, O.J., Hoffmeister, J.K., Coates, C., and White, B.J. A partial evaluation of Torrance's tests of creativity. Amer. educ. Res. J. 7 (197o), S. 359 - 372

Hasan, P., and Butcher, H.J. Creativity and intelligence : a partial replication with Scottish children of Getzels' and Jackson's study. Brit. J. Psychol. 57 (1966) 1-2, S. 129 - 135

Hatcher, C., Felker, D.W., and Treffinger, D.J. The prediction of upper grade reading achievement with measures of intelligence, divergent thinking, and self-concept. Paper presented at the 58 th annual meeting of the American Educational Research Association, Chicago, April 1974

Hauck, W.E., and Thomas, J.W. The relationship of humor to intelligence, creativity, and intentional and incidental learning. J. exp. Educ. 4o (1972) 4, S. 52 - 55

Heinelt, G. Kreative Lehrer - kreative Schüler. Freiburg : Herder 1974

Heller, K. und Rosemann, B. Planung und Auswertung empirischer Untersuchungen. Stuttgart: Klett 1974

Herbig, M., und Hollensteiner, P. Eine Unterrichtssequenz zur Förderung kreativen Verhaltens und die Überprüfung ihrer Effektivität. Die Deutsche Schule 65 (1973) 4, S. 224 - 235

Herrmann, T. und Stapf, K.H. Über theoretische Konstruktionen in der Psychologie. In: Berichte aus dem Institut für Psychologie der Philipps-Universität in Marburg/Lahn, Nr. 24, 197o

Hofer, M. Die Schülerpersönlichkeit im Urteil des Lehrers. Weinheim : Beltz 1969

Holland, J.L. Some limitations of teacher ratings as predictors of creativity. J. educ. Psychol. 5o (1959) 5, S. 219 - 223

Holland, J.L. E.P. Torrance. Torrance Tests of Creative Thinking. J. counsel. Psychol. 15 (1968) 3, S. 297 - 298

Holland, J.L., and Richards, J.M., Jr. The many faces of talent: a reply to Werts. J. educ. Psychol. 58 (1967) 4, S. 2o5-2o9

Hollensteiner, P. Eine Trainingsmethode und ihr Einfluß auf die Kreativitätsleistung von Schülern einer 7. Klasse. Kunst und Unterricht 1974, Heft 23 (Februar), S. 28-33

Horn, J.L. On subjectivity in factor analysis. In: Educ. psychol. Measmt. 27 (1967), S. 811-82o

Horn, J.L., and Knapp, J.R. On the subjective character of the empirical base of Guilford's Structure-of-Intellect Model. Psychol. Bull. 8o (1973) 1, S. 33 - 43

Horn, R. Sozialschicht und Intelligenzleistung. Psychologie in Erziehung und Unterricht 21 (1974) 1, S. 58-62

Horn, W. Prüfsystem für Schul- und Bildungsberatung P-S-B. Göttingen : Hogrefe 1969

Ingenkamp, K., Wolf, B. et al. Bildungs-Beratungs-Test für 4. - 6. Klassen BBT 4 - 6. Weinheim : Beltz 1977

Ingersoll, G.M., and Snowman, J. A two year evaluation of the test - retest stability of the Torrance Tests of Creative Thinking with fifth and sixth grade children. Paper presented to the Annual Meeting of the American Educational Research Association, Chicago, 1974

Janis, I.L. (Ed.), Mahl, G.F., Kagan, J., and Holt, R.R. Personality. Dynamics, development, and assessment. New York : Harcourt, Brace and World 1969, S. 475 - 491

Kaiser, H.F. Computer program for varimax rotation in factor analysis. Educ. psychol. Measmt. 19 (1959), S. 413-42o

Kaltsounis, B. Instruments useful in studying creative behavior and creative talent. J. creat. Behav. 5 (1971) 2, S. 117 - 126

Kaminski, G. Verhaltenstheorie und Verhaltensmodifikation. Stuttgart : Klett 197o

Kaplar, J.E. Creativity, sex-role preference, and perception of parents in fifth-grade boys. Diss. Abstr. 197o, 3o-B, S. 5689

Kasielke, E. Zur Diagnostik des sprachlichen Entwicklungsstandes bei Vorschulkindern. Z. Psychol. mit Z. angew. Psychol. 174 (1967), S. 1 - 67

Keenan, J.F. The relationship of certain socio-cultural and community factors among sixth grade students to creativity in art. Diss.Abstr. 1971, 31 8-A, S. 3782

Kemmler, L. Erfolg und Versagen in der Grundschule. Göttingen : Hogrefe 197o[2]

Kemmler, L. und Langheinrich, D. Eine deutsche Fassung der "Primary Mental Abilities " für acht-bis zehnjährige Kinder. In : Bericht über den 25. Kongreß der Deutschen Gesellschaft für Psychologie. Göttingen : Hogrefe 1967, S. 669 - 673

Kerlinger, F.N., and Pedhazur, E.J. Multiple regression in behavioral research. New York : Holt, Rinehart and Winston 1973

Klausmeier, H.J., Harris, C.W., and Ethnathios, Z. Relationships between divergent thinking abilities and teacher ratings of high school students. J. educ. Psychol. 53 (1962) 2, S. 72-75

Kleining, G. und Moore, H. Soziale Selbsteinstufung (SSE). Ein Instrument zur Messung sozialer Schichten. Köln. Z. Soziol. Sozialpsychol. 2o (1968) 3, S. 5o2-552

Knoche, W. Jungen, Mädchen, Lehrer und Schulen im Zensurenvergleich. Weinheim : Beltz 1969

Kogan, N. A clarification of Cropley and Maslany's analysis of the Wallach-Kogan creativity tests. Brit. J. Psychol. 62 (1971) 1, S. 113 - 117

Kogan, N. Creativity and sex differences. J. creat. Behav. 8 (1974) 1, S. 1 - 14

Kogan, N., and Morgan, F.T. Task and motivational influences on the assessment of creative and intellective ability in children. Genet. Psychol. Monogr. 8o (1969), S. 91-127

Kogan, N., and Pankove, E. Creative ability over a five-year span. Research Bulletin des Educational Testing Service, Princeton, N.J., September 1971

Krause, R. Kreativität. Untersuchungen zu einem problematischen Konzept. München : Goldmann 1972

Krüger, K., Hylla, E. und Bargemann, R. Zahlenrechnen ZR 4 +. Aus der Reihe " Deutsche Schultests ". Weinheim : Beltz 197o

Landau, E. Psychologie der Kreativität. München: Ernst Reinhardt 1971[2]

Laux, L. und Zinn, A. Auswirkungen stressinduzierender Situationsfaktoren auf divergentes Problemlösen. Vortrag, gehalten auf dem 29. Kongreß der Deutschen Gesellschaft für Psychologie in Salzburg, 1974

Leith, G. The relationships between intelligence, personality and creativity under two conditions of stress. Brit. J. educ. Psychol. 42 (1972), S. 24o - 247

Lienert, G.A. Testaufbau und Testanalyse. Weinheim:Beltz 1967[2]

Lohmann, J. Kreativität, Persönlichkeit, Erziehung. Unveröffentlichte Dissertation des Fachbereichs I der Universität in Trier. 1975

Long, B.H., and Henderson, E.H. Opinion formation and creativity in elementary school children. Psychol. Rep. 17 (1965) 1, S. 219-223

Lückert, H.R. Die Ergebnisse der Kreativitätsforschung. In: Lückert, H.R. (Hrsg.) Begabungsforschung und Bildungsförderung als Gegenwartsaufgabe. München: Ernst Reinhardt 1969, S. 128-144

Lytton, H.,and Cotton, A.C. Divergent thinking abilities in secondary schools. Brit. J. educ. Psychol. 39 (1969), S. 188 - 19o

Mac Kinnon, D.W. The nature and nurture of creative talent. Amer. Psychologist 17 (1962), S. 484-495

Mac Kinnon, D.W. The study of creative persons : a method and some results. In : Kagan, J. (Ed.) Creativity and learning. Boston 1967, S. 2o-35

Mackler, B., and Shontz, F.W. Characteristics of responses to tests of creativity. J. clin. Psychol. 23 (1967), S. 73-8o

Madaus, G. F. Divergent thinking and intelligence. Another look at a controversial question. J. educ. Measmt. 4 (1967) 4, S. 227 - 235

Mainberger, U. Kreativitätsförderung in der Schule. Neue Sammlung 15 (1975) 3, S. 231 - 252

Marsh, R.W. Research note. A statistical re-analysis of Getzels and Jackson's data. Brit. J. educ. Psychol. 34 (1964), S. 91-93

Massad, C.E. A comparative study of creativity, language aptitude, and intelligence in sixth-grade children from low-socioeconomic and middle-socioeconomic levels. Diss. Abstr. 1969, 29, 12, 1-A, S. 4331

Massialas, B.E. und Zevin, J. Kreativität im Unterricht. Stuttgart : Ernst Klett 1969

Matussek, P. Kreativität als Chance. München : Piper 1974

Mc Cutcheon, N.S. A study of relationship among creativity, intelligence,and test anxiety of middle-class fourth grade boys and girls.Diss. Abstr. 197o, 3o, 11 - A, S. 4833

Mc Kinney, J.D., and Forman, S.G. Factor structure of the Wallach-Kogan tests of creativity and measures of intelligence and achievement. Psychology in the Schools 14 (1977) 1, S. 41-44

Mc Nemar, Q. Lost: our intelligence? Why? Amer. Psychologist 19 (1964), S. 871 - 882

Mc Whinnie, H.J. Some relationships between creativity and perception in sixth-grade children. Percep. mot. Skills 25 (1967), S. 979 - 98o

Mc Whinnie, H.J. Some relationships between creativity and perception in fourth children. Acta Psychol. 31 (1969), S. 169 - 175

Mc Whinnie, H.J. A factor analytic study of perceptual behavior in 4 th and 5 th grade children. Acta Psychol. 34 (197o), S. 89 - 97

Mc Whinnie, H.J. Some relationships between creativity and perception in sixth grade children. Scientia Paedagogica Experimentalis 5 (1968) 1, S. 84 - 9o

Mednick, S.A. The associative basis of the creative process. Psychol. Rev. 69 (1962) 3, S. 22o - 232

Mednick, S.A., and Mednick, M.T. An associative interpretation of the creative process. In : Taylor, C.W. (Ed.) Widening horizons in creativity. New York : Wiley 1964, S. 54 - 79

Messick, S. Multivariate models of cognition and personality: the need for both process and structure in psychological theory and measurement. Research Bulletin des Educational Testing Service, Princeton, N.J., August 1973

Michel, L. Allgemeine Grundlagen psychometrischer Tests. In : Heiss, R. (Hrsg.) Handbuch der Psychologie, 6. Band, Psychologische Diagnostik. Göttingen : Hogrefe 1964, S. 19 - 7o

Mollenhauer, K. Sozialisation und Schulerfolg. In : Roth, H. (Hrsg.) Begabung und Lernen. Stuttgart : Klett 1971[6], S. 269 - 296

Mondfrans, A. P. van, Feldhusen, J.F., Treffinger, D.J., and Ferris, D.R. The effects of instructions and response time on divergent thinking test scores. Psychology in the Schools 8 (1971) 1, S. 65 - 71

Mühle, G., und Schell, C. (Hrsg.) Kreativität und Schule. München: Piper 1970

Murphy, R.T. Investigation of a creativity dimension. Educational Testing Service, Princeton, N.J., February 1973. (Unveröffentlichte Dissertation)

Mutschler,D. Intelligenz und Kreativität. Z. f. Päd. 15 (1969) 2, S. 119 - 133

Nicholls, J.G. Wallach, M.A., and Wing, C.W. The talented student. A validation of the creativity - intelligence distinction. J. gen. Psychol. 117 (197o), S. 277 - 279

Nicholls, J.G. Some effects of testing procedure on divergent thinking. Child Developm. 42 (1971), S. 1647 - 1651

Nicholls, J.G. Creativity in the person who will never produce anything original and useful:the concept of creativity as a normally distributed trait. Amer. Psychologist 27 (1972) 2, S. 717 - 727

Nichols, R.C. Parental attitudes of mothers of intelligent adolescents and creativity of their children. Child Developm. 35 (1964), S. 1o41-1o49

O'Bryan, K.G., and Mc Arthur, R.S. Reversibility, intelligence, and creativity in nine-year-old boys. Child Developm. 4o (1969), S. 33 - 45

Oevermann, U. Schichtenspezifische Formen des Sprachverhaltens und ihr Einfluß auf die kognitiven Prozesse. In : Roth, H. (Hrsg.) Begabung und Lernen. Stuttgart : Klett 1971, S. 297 - 356

Ogletree, E. A cross-cultural examination of the creative thinking ability of public and private school pupils in England, Scotland, and Germany. J. soc. Psychol. 83 (1971), S. 3o1 - 3o2

Ogletree, E.J., and Ujlaki, W. Effects of social class status on tests of creative behavior. J. educ. Res. 67 (1973) 4, S. 149-152

Phillips, V.K.,and Hudgins, Q.L. Relationship between creativity, sex, and rule-orientation behavior. Percept. mot. Skills 38 (1974), S. 1163 - 1171

Piers, E.V., Daniels, J.M., and Quackenbush, J.F. The identification of creativity in adolescents. J. educ. Psychol. 51 (196o) 6, S. 346 - 351

Porter, R.B. A comparative investigation of the personality of sixth-grade gifted children and a norm group of children. J. educ. Res. 58 (1964) 3, S. 132-134

Porter, R.B., and Cattell, R.B. Children's Personality Questionnaire " The CPQ ". Champaign, Illinois: Institute for Personality and Ability Testing 1963

Rahm, D. Untersuchung über den Zusammenhang von repressiver Erziehungseinstellung und Kreativität. Psychologie in Erziehung und Unterricht 21 (1974) 5, S. 259 - 269

Richards, R.L. A method-dependent resolution of Guilford and Wallach-Kogan descriptions of creative thinking. Unpublished paper 1974 (Adresse: Boston University, School of Education, Department of Foundations. 765 Commonwealth Avenue. Boston, Massachusetts o2215)

Ripple, R.E., and May, F.B. Caution in comparing creativity and IQ. Psychol. Rep. 1o (1962), S. 229-23o

Rookey, T.J. Creativity:Concepts and explorations. Paper presented to the American Educational Research Association, Symposion in New Orleans, Spring 1973

Rookey, T.J., and Reardon, F.J. The interdependence of creative attitude and creative ability. Paper presented to the American Educational Research Association, Symposion in New Orleans, Spring 1973

Rotter, D.M., Langland, L., and Berger, D. The validity of tests of creative thinking in seven-year-old children. Gifted Child Quart. 15 (1971) 4, S. 273-278

Schaefer, C.E., and Anastasi,A. A biographical inventory for identifying creativity in adolescent boys. J. appl. Psychol. 52 (1968) 1, S. 42 - 48

Schiffer, H. Fragen zur Kreativität. Workshop Schulpädagogik, Materialien 6. Ravensburg : Otto Maier 1973

Schreiner, G. Kreativität und Anpassung. Die Deutsche Schule 63 (1971), S. 19 - 31

Schröder, H. Kombinierter Lern- und Intelligenztest KLI 4 +. Hrsg.: K. Ingenkamp. Aus der Reihe "Deutsche Schultests" Weinheim : Beltz 1968

Schulte, D. Feldabhängigkeit in der Wahrnehmung. Meisenheim am Glan : Anton Hain 1974

Seiffge-Krenke, I. Probleme und Ergebnisse der Kreativitäts-forschung. Bern: Hans Huber 1974

Seisenberger, G. Kreativität und Lernerfolg. Welt der Schule 27 (1974) 8, S. 281 - 291

Siegelman , M. Parent behavior correlates of personality traits related to creativity in sons and daughters. J. consult. clin. Psychol. 4o (1973) 1, S. 43-47

Solomon, A.O. A comparative analysis of creative and intelligence behavior of elementary school children with different socioeconomic backgrounds. Diss. Abstr. 1968, 5-A, S. 1457

Spotts, J.V., and Mackler, B. Relationships of field-dependent and field-independent cognitive styles to creative test performance. Percept. mot. Skills 24 (1967), S. 239 - 268

Stammeier, E. Kreativität im Unterricht? Lebendige Schule 29 (1974) 9, S. 326 - 334

Stark, G., Bauer, A. et al. Theorie und Praxis der Leistungs-messung in Gesamtschulen. Projektgruppe Leistungs-messung in Gesamtschulen am Deutschen Institut für Internationale Pädagogische Forschung. Frankfurt/Main 1973

Stark, W. Die Sitzenbleiber-Katastrophe. Stuttgart : Klett 1974

Stevens, J.C. A study of the relationships between field independence, dogmatism, and creativity in rural seventh grade students. Diss. Abstr. 197o, 3o, 7-A, S. 2863

Stubbings, J.R., Jr. A comparison of the Torrance Tests of Creative Thinking and Guilford's measures of creative ability on sex, cognitive and personality variables. Diss. Abstr. 1968, 28, 11-A, S. 4496

Sullivan, J. The relationship of creative and convergent thinking to literal and critical reading ability of children in the upper grades. J.educ. Res. 66 (1973) 8, S. 374 - 377

Taylor, C.W., and Barron, F. (Eds.) Scientific creativity : its recognition and development. New York : Wiley 1963

Thorndike, R.L. Some methodological issues in the study of creativity. In: Proceedings of the invitational conference on testing problems. Princeton, N.J. : Educational Testing Service 1962, S. 4o - 54

Thurner, F. und Tewes, U. Kinder-Angst-Test K-A-T. Göttingen : Hogrefe 1969

Torrance, E.P. Give the " Devil " his dues. Gifted Child Quarterly 5 (1961), S. 115 - 118

Torrance, E.P. Guiding creative talent. Englewood Cliffs, N.J.:Prentice-Hall 1962 a

Torrance, E.P. Cultural discontinuities and the development of originality of thinking . Exceptional Children 1962b (September), S. 2 - 13

Torrance, E.P. Education and the creative potential. Minneapolis : University of Minnesota Press 1963 a

Torrance, E.P. Cultural discontinuities and the development of originality. In : Torrance, E.P. Education and the creative potential. Minneapolis : University of Minnesota Press 1963 b

Torrance, E.P. The Minnesota studies of creative thinking : 1959 - 62. In : Taylor, C.W. (Ed.) Widening horizons in creativity. New York : Wiley 1964 a, S. 125 - 143

Torrance, E.P. Creative teaching makes a difference. The Florence S. Dunlop Memorial Lecture, Ontario Council for Exceptional Children, Ontario 1964 b

Torrance, E.P. Rewarding creative behavior. Englewood Cliffs, N.J.:Prentice-Hall 1965a

Torrance, E.P. Gifted children in the classroom. New York 1965b

Torrance, E.P. Torrance Tests of Creative Thinking : Norms-technical manual. Princeton, N.J.: Personnel Press 1966

Torrance, E.P. The Minnesota studies of creative behavior: national and international extensions. J. creat. Behav. 1 (1967a) 2, S. 137 - 154

Torrance, E.P. Nurture of creative talents. In : Mooney, R.L., and Razik, T.A. (Eds.) Explorations in creativity. New York : Harper and Row 1967 b.

Torrance, E.P. Neue Item-Arten zur Erfassung kreativer Denkfähigkeiten. In : Ingenkamp, K. und Marsolek, Th. (Hrsg.) Möglichkeiten und Grenzen der Testanwendung in der Schule. Weinheim : Beltz 1968 a, S. 325 - 343

Torrance, E.P. Examples and rationales of test tasks for assessing creative abilities. J. creat. Behav. 2 (1968 b) 3, S. 165 - 178

Torrance, E.P. Curiosity of gifted children and performance on timed and untimed tests of creativity. Gifted Child Quart. 13 (1969) 3, S. 155-158

Torrance, E.P. Achieving socialization without sacrificing creativity. J. creat. Behav. 4 (197o) 3, S. 183-189

Torrance, E.P. Identity : the gifted child's major problem. Gifted Child Quarterly 15 (1971 a) 3, S. 147 - 155

Torrance, E.P. Are the Torrance Tests of Creative Thinking biased against or in favor of " disadvantaged " groups? Gifted Child Quart. 15 (1971 b) 2, S. 75-79

Torrance, E.P. Can we teach children to think creatively? J. creat. Behav. 6 (1972a) 2, S. 114 - 143

Torrance, E.P. Predictive validity of the Torrance Tests of Creative Thinking. J. creat.Behav. 6 (1972 b) 4, S. 236 - 252

Torrance, E.P. Interscholastic brainstorming and creative problem solving competition for the creatively gifted. Gifted Child Quarterly 18 (1974) 1, S. 3 -7

Torrance, E.P., Tan, C.A., and Allman, T. Verbal originality and teacher behavior : a predictive validity study. J. of. Teacher Education 21 (197o) 3, S. 335 - 341

Ulmann, G. Kreativität. Weinheim : Beltz 1968

Vernon, P.E. Creativity and intelligence. Educ. Res. 6 (1964) 3, S. 163 - 169

Vernon, P.E. Ability factors and environmental influences. In : Wolfle, D. (Ed.) The discovery of talent. Cambridge, Mass.: Harvard University Press 1969, S. 279 - 295

Vernon, P.E. Effects of administration and scoring on divergent thinking tests. J. educ. Psychol. 41 (1971) 3, S. 245 - 257

Viebahn, P. Zur Kreativität als Problem-Prozeß-Produkt-Einheit. Welt der Schule 23 (197o) 12, S. 445-449

Wagner, S.B. Timed and untimed measures of creativity and their correlates. Diss.Abstr. Internat. 1972, 33, 5-A, S. 2185

Wallach, M.A. Creativity. In : Mussen, P.H. (Ed.) Carmichael's manual of child psychology. Volume 1. New York : Wiley 197o^3 , S. 1211-1272

Wallach, M.A.,and Kogan,N. Modes of thinking in young children. New York : Holt, Rinehart and Winston 1965

Wallach, M.A., and Wing, C.W., Jr. The talented student. A validation of the creativity-intelligence distinction. New York : Holt, Rinehart and Winston 1969

Ward, W.C. Creativity in young children. Child Developm. 39 (1968), S. 737 - 754

Ward, W.C. Rate and uniqueness in children's creative responding. Proceedings of the 77 th Annual Convention of the American Psychological Association 4 (1969a) 1, S. 251 - 252

Ward, W.C. Creativity and environmental cues in nursery school children. Development. Psychol. 1 (1969 b) 5, S. 543 - 547

Ward, W.C. Convergent and divergent measurement of creativity in children.Research Bulletin des Educational Testing Service, Princeton, N.J., June 1971

Ward, W.C. A field study of nonverbal creativity. Research Bulletin des Educational Testing Service, Princeton, N.J. August 1971

Ward, W.C., Kogan, N., and Pankove, E. Incentive effects in children's creativity. Child Developm. 43 (1972) 2, S. 669 - 676

Wasik, J.L. Teacher perceptions of behaviors associated with creative problem solving performance. Educ. psychol. Measmt. 34 (1974)2, S. 327 - 341

Weisberg, P.S., and Springer, K.J. Environmental factors in creative function. Archives of General Psychiatry 5 (1961), S. 554 - 564

Weiß, R. Die Zuverlässigkeit der Ziffernbenotung bei Aufsätzen und Rechenarbeiten. In : Ingenkamp, K. (Hrsg.) Die Fragwürdigkeit der Zensurengebung. Weinheim : Beltz 1971, S. 9o - 1o2

Werner, E.E. CPQ personality factors of talented and under-achieving boys and girls in elementary school. J. clin. Psychol. 22 (1966), S. 461 - 464

Werner, E.E.,and Bachtold, L.M. Personality factors of gifted boys and girls in middle childhood and adolescence. Psychology in the Schools 6 (1969) 2, S. 177 - 182

Werts, C.E. The many faces of intelligence. J. educ. Psychol. 58 (1967) 4, S. 198 - 2o4

Wewetzer, K.H. Intelligenz und Intelligenzmessung. Darmstadt : Wissenschaftliche Buchgesellschaft 1972

Wieczerkowski, W., Nickel, H., Janowski, A., Fittkau, B. und Rauer, W. Angstfragebogen für Schüler (AFS). Braunschweig : Westermann 1974

Williams, F.E. Modelle zur Förderung der Kreativität in der Klasse durch Integration von kognitivem und affektivem Verhalten. Neue Unterrichtspraxis 1975,5, S. 3o3 - 31o

Williams, T.M, and Fleming, J.W. Methodological study of the relationship between associative fluency and intelligence. Development. Psychol. 1 (1969) 2, S. 155 - 162

Wilson, R.C., Guilford, J.P., Christensen, P.R., and Lewis, D.J. A factor-analytic study of creative-thinking abilities. Psychometrika 19 (1954) 4, S. 297 - 311

Witkin, H.A. The role of cognitive style in academic performance and in teacher-student relations. Educational Testing Service, Princeton, N.J., February 1973

Witkin, H.A., Price-Williams, D., et al. Social conformity and psychological differentiation. Research Bulletin des Educational Testing Service, Princeton, N.J., October 1973

Wodtke, K.H. Some data on the reliability and validity of creativity tests at the elementary school level. Educ. psychol. Measmt. 24 (1964) 2, S. 399-4o8

Wolf, V.C. Age and sex performance differences as measured by a new nonverbal visual perceptual test. Psychonomic Science 25 (1971) 2, S. 85 - 87

Wollschläger, G. Kreativität und Gesellschaft. Frankfurt : Fischer Taschenbuch 1971

Yamamoto, K. Creativity and sociometric choice among adolescents. J. soc. Psychol. 64 (1964 a), S. 249-261

Yamamoto, K. Role of creative thinking and intelligence in high school achievement. Psychol. Rep. 14 (1964 b), S. 783 - 789

Yamamoto, K. Research frontier. " Creativity " -a blind man's report on the elephant. J. counsel. Psychol. 12 (1965 a) 4, S. 428 - 434

Yamamoto, K. Effects of restriction of range and test unreliability on correlation between measures of intelligence and creative thinking. Brit. J. educ. Psychol. 35 (1965 b), S. 3oo - 3o5

Yamamoto, K. Creativity and unpredictability in school achievement. J. educ. Res. 6o (1967 a) 7, S. 321-325

Yamamoto, K. Validation of tests of creative thinking : a review of some studies. In : Mooney, R., and Razik, T.A. (Eds.) Explorations in creativity. New York : Harper and Row 1967 b

Yamamoto, K., and Chimbidis, M.E. Achievement, intelligence, and creative thinking in fifth-grade children : a correlational study. Merrill-Palmer Quart. 12 (1966), S. 233 - 241

Anhang A

A1

Erste und zweite Versuchsform (II,4 und II,5)

Erziehungswissenschaftliche Hochschule Rheinland-Pfalz
Zentrum für empirische pädagogische Forschung

Versuchsform II,4

Du hast sicher schon an einem oder sogar an mehreren Tests teilgenommen. Erinnerst Du Dich noch daran? Sicher hast Du bei einigen der Aufgaben ein bißchen Angst gehabt, daß Du sie nicht lösen kannst. Auf jeden Fall war es fast immer so, daß Deine Lösungen entweder richtig oder falsch waren.
Heute habe ich einige Aufgaben mitgebracht, für die es keine falschen Antworten gibt. Diese Aufgaben kann man nur mehr oder weniger einfallsreich lösen. Besonders gut kannst Du sie lösen, wenn Du Deiner Phantasie freien Lauf läßt, Dir möglichst viele Gedanken einfallen läßt und auch ausgefallene Antworten gibst, an die vielleicht außer Dir niemand denkt. Schau also nicht nach, was Dein Nachbar schreibt oder zeichnet, sondern denk Dir selbst etwas aus! Übrigens, die Schrift und die Rechtschreibung sind heute einmal nicht so wichtig. Die Hauptsache ist, daß wir hinterher wissen, was Du gemeint hast.
Du wirst sehen, daß recht lustige Aufgaben dabei sind, die Dir sicher Spaß machen werden. Wenn alle fertig sind, erzählen wir uns dann unsere lustigsten Einfälle oder wir malen sie an die Tafel.

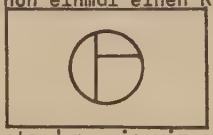
Name: _____ Vorname: _____

Schule: _____ Klasse: _____

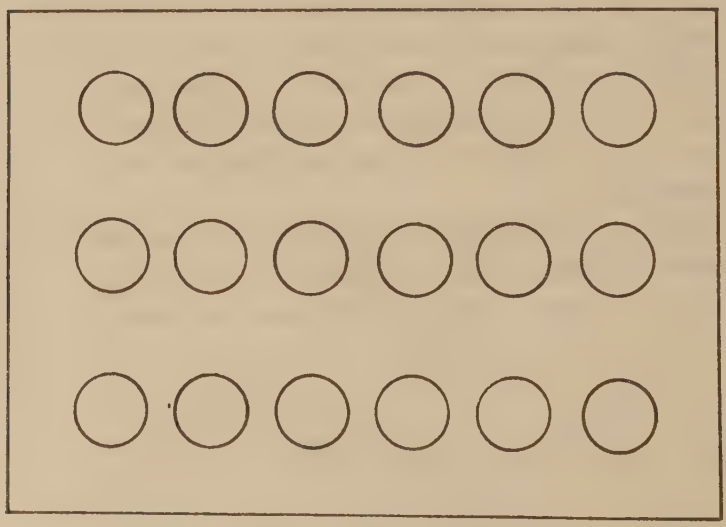
1. Kreise teilen

Jetzt kommt ein Spiel, bei dem Du die gleichen Kreise auf möglichst viele Arten in 3 Stücke zu teilen hast.
Dabei können die Stücke verschieden groß sein, aber es müssen immer 3 Stücke sein. Beachte, daß die Stücke durch gerade Linien voneinander zu trennen sind.
Wir wollen nun einmal einen Kreis teilen:

Beispiel:



Hast Du verstanden, wie wir es machen wollen?
Du sollst möglichst viele Kreise auf verschiedene Arten in 3 Stücke teilen.



Bitte nicht umblättern!

Eigenschaften

Schreibe möglichst viele Gegenstände auf, die folgende unterstrichenen Eigenschaften haben.

2. Schreibe alle Gegenstände auf, die leuchten.

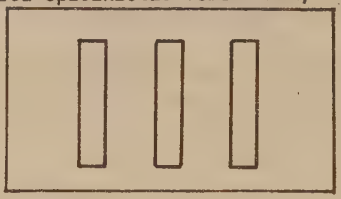
3. Schreibe alle Gegenstände auf, die sich drehen.

4. Schreibe alle Gegenstände auf, die zerbrechen können.

Bitte nicht umblättern!

5. Spielklötze

In diesem Spiel sollst Du Dir drei Spielklötze vorstellen, so wie sie hier gezeichnet sind.



Mit diesen Klötzen sollst Du in Gedanken alle Figuren bauen, die Dir einfallen. Zeichne dann die Figuren auf diese Seite. Die Linien auf dem Papier helfen Dir, damit Du die Figuren besser zeichnen kannst und kein Lineal verwenden mußt. Die Klötze sollen nicht umfallen. Zeichne die Figuren auch so, daß die Klötze stehen bleiben.
Versuche nun, möglichst viele verschiedene Figuren zu zeichnen.

Bitte nicht umblättern!

6. Zahlensalat

Du hast hier eine lange Reihe von Zahlen, die durcheinander gepurzelt sind.

9 7 4 5 0 2 6 1 3 8 2 7 0 3 4 1 5

Du sollst jetzt ein bißchen Ordnung in die Zahlen bringen. Bilde deshalb möglichst viele Gruppen von solchen Zahlen, die irgendwie zusammenpassen. Eine Gruppe muß aber aus mindestens 4 Zahlen bestehen.
Hier sind 2 Beispiele:

1 2 3 4

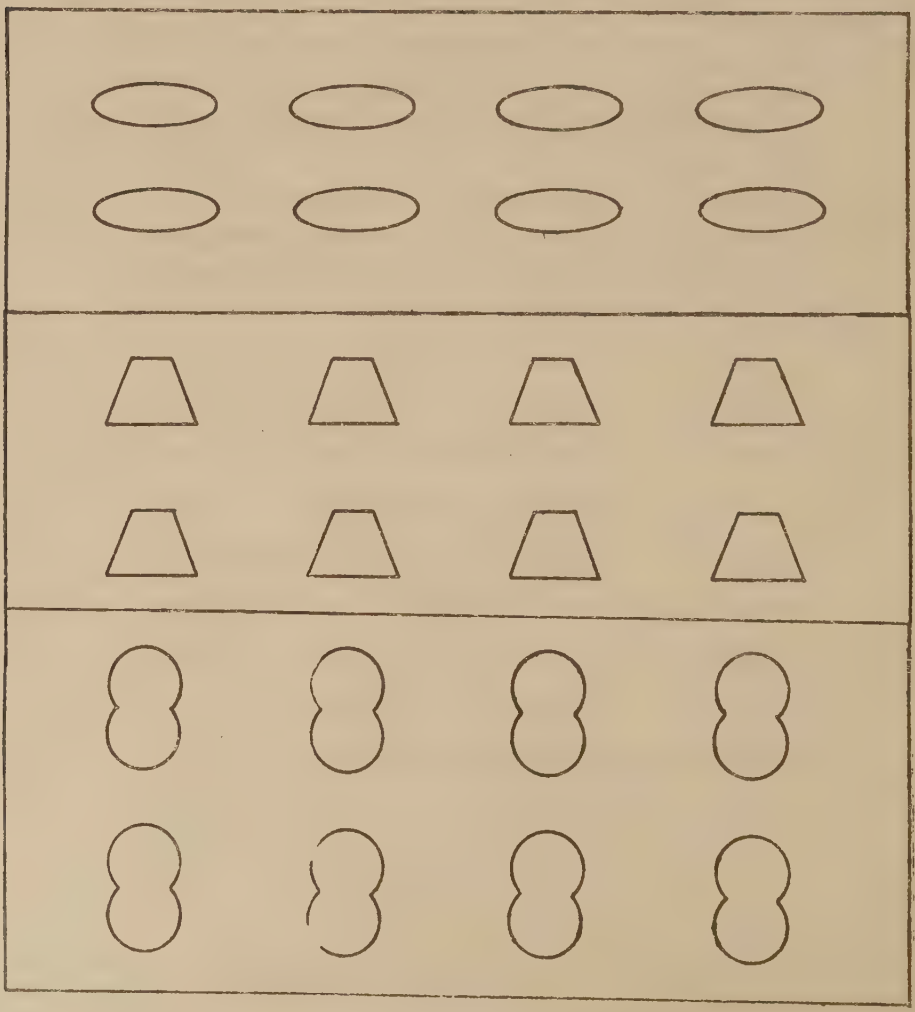
5 3 1 3 5

In beiden Reihen sind die Zahlen auf eine bestimmte Weise geordnet, sie bilden eine Gruppe. Denke Dir jetzt möglichst viele solcher Gruppen aus und schreibe sie auf.

Bitte nicht umblättern!

7. Zeichnungen

Nun kommen wir zum nächsten Spiel. Du siehst hier immer die gleichen Figuren. Zeichne daraus bekannte Gegenstände, in dem Du andere Zeichen oder Striche hinzufügst. Es kommt jedoch nicht darauf an, daß Du sehr schön zeichnest, sondern es ist wichtig, daß Dir möglichst viele Gegenstände einfallen. Damit wir erkennen können, was Du gemeint hast, schreibe zu jeder Zeichnung dazu, was es sein soll.



Verrückte Geschichten

Jetzt wollen wir uns einmal ganz verrückte Fragen überlegen. Stell Dir einmal vor, es würde keinen Winter mehr geben. Was würde dann alles geschehen? Man könnte z.B. keinen Schlitten fahren, man brauchte keinen Pelzmantel usw.

8. Nun stelle Dir einmal vor, alle Menschen würden sich von morgen an nur noch mit 3 Tabletten am Tag ernähren. Was würde sich daraus alles ergeben?
   Schreibe alles auf, was Dir dazu einfällt.

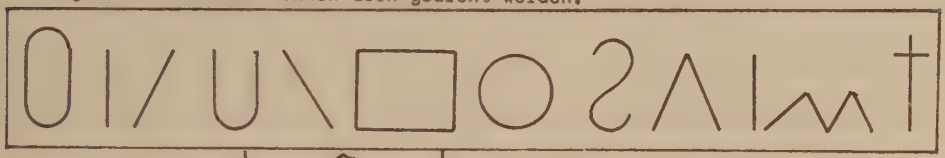
9. Nimm einmal an, alle Ameisen wären so groß, wie die Menschen. Was könnte sich daraus ergeben?

10. Nimm einmal an, Du könntest frei in der Luft schweben.
Was könnte da alles passieren?

Bitte nicht umblättern!

11. Konstruktionen

Bei diesem Spiel sollst Du aus einer Reihe von Zeichen und Figuren möglichst viele einfache Gegenstände zeichnen, die etwas bedeuten. Du brauchst dabei nicht alle Zeichen zu verwenden, sondern nur diejenigen, die Du für den jeweiligen Gegenstand benötigst. Jedoch darfst Du kein anderes Zeichen hinzufügen. Die Zeichen können auch gedreht werden.



Beispiel:

Dieses Haus entsteht, wenn man das Viereck und zwei Striche zusammenfügt. Zeichne nun möglichst viele verschiedene Gegenstände.

Bitte nicht umblättern!

12. Schulfrei

Stell Dir vor, Ihr hättet statt eines Wandertages einen unterrichtsfreien Tag, den die Schüler selbst gestalten dürfen. Dieses Mal hast Du den Auftrag erhalten, einen Plan für diesen Tag zu machen. Euer Lehrer hat nur eine Bedingung gestellt, nämlich daß Ihr abends, bevor es dunkel wird, wieder zu Hause seid.
Schreibe in Stichworten, welche Vorschläge Du machen würdest.

Bitte nicht umblättern!

13. Buchstabenspiel

Jetzt kommt ein Spiel, in dem Du Deiner Phantasie wieder freien Lauf lassen kannst.
Du sollst diesmal viele Möglichkeiten finden, wie man denselben Buchstaben schreiben, malen oder zeichnen kann. Versuche dies einmal mit dem Buchstaben O. Zeichne die Buchstaben aber so, daß man sie noch als O erkennen kann.

Bitte nicht umblättern!

Sätze machen

Nun machen wir ein neues Spiel. Du siehst hier immer vier Anfangsbuchstaben von Wörtern, die klein oder groß geschrieben sein können. Versuche aus diesen Buchstaben möglichst viele Vier-Wort-Sätze zu bilden. Du kannst die Buchstaben auch in einer anderen Reihenfolge verwenden.

Beispiel:	E	G	Z	S
	Edgar	geht	zur	Schule.
	Gabi	schält	eine	Zwiebel.

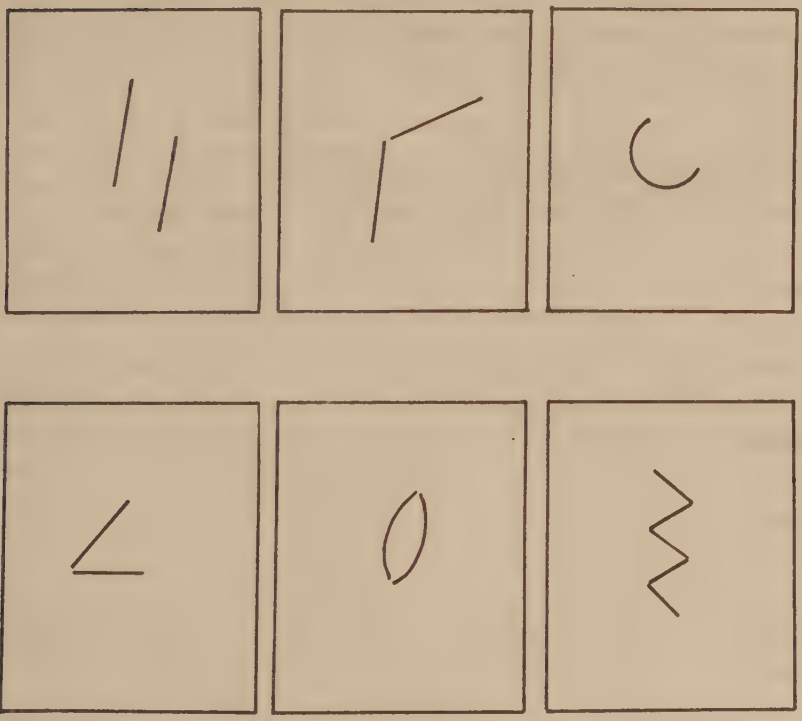
Beim ersten Satz stehen die Anfangsbuchstaben in der angegebenen Reihenfolge; im zweiten Satz wurde die Reihenfolge verändert.
Nun mache aus den folgenden Anfangsbuchstaben möglichst viele Sätze.

14. H D S G

15. T F I B

16. Unvollständige Figuren

Du findest jetzt 6 verschiedene Zeichen. Versuche aus jedem dieser Zeichen einfache Bilder oder Figuren zu zeichnen. Dabei mußt Du zu den vorgegebenen Zeichen Linien hinzufügen. Zeichne solche Figuren, an die niemand sonst denken wird, und schreibe jeweils darunter, was Du gezeichnet hast.

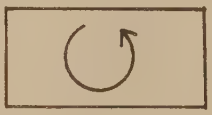


Bitte nicht umblättern!

17. Zeichen erfinden

Nun kannst Du einmal zeigen, ob Du erfinderisch bist. Es werden Dir jetzt einige Gegenstände und Tätigkeiten vorgegeben, zu denen Du einfache Zeichen erfinden sollst.
So könntest Du z.B. für das Wort 'Auto' folgendes verwenden:

Dieses Zeichen stellt ein Lenkrad dar. Da das Lenkrad ein sehr wichtiger Teil des Autos ist, kann sich jeder gleich vorstellen, daß mit diesem Zeichen ein Auto gemeint ist. Für das Wort'drehen' könnte man z.B. dieses Zeichen setzen:



Dies ist ein Zeichen dafür, daß sich etwas im Kreis dreht.
Erfinde nun für die folgenden Wörter möglichst einfache Zeichen.

Radio

springen

Spielplatz

trinken

Küche

fliegen

Schule

lesen

Erziehungswissenschaftliche Hochschule Rheinland-Pfalz

Zentrum für empirische pädagogische Forschung

Versuchsform II,5

Du hast sicher schon an einem oder sogar an mehreren Tests teilgenommen. Erinnerst Du Dich noch daran ? Sicher hast Du bei einigen der Aufgaben ein bißchen Angst gehabt, daß Du sie nicht lösen kannst. Auf jeden Fall war es fast immer so, daß Deine Lösungen entweder richtig oder falsch waren.
Heute habe ich einige Aufgaben mitgebracht, für die es keine falschen Antworten gibt. Diese Aufgaben kann man nur mehr oder weniger einfallsreich lösen. Besonders gut kannst Du sie lösen, wenn Du Deiner Phantasie freien Lauf läßt, Dir möglichst viele Gedanken einfallen läßt und auch ausgefallene Antworten gibst, an die vielleicht außer Dir niemand denkt. Schau also nicht nach, was Dein Nachbar schreibt oder zeichnet, sondern denk Dir selbst etwas aus! Übrigens, die Schrift und die Rechtschreibung sind heute einmal nicht so wichtig. Die Hauptsache ist, daß wir hinterher wissen, was Du gemeint hast.

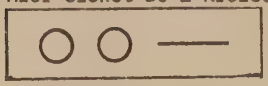
Du wirst sehen, daß recht lustige Aufgaben dabei sind, die Dir sicher Spaß machen werden. Wenn alle fertig sind, erzählen wir uns dann unsere lustigsten Einfälle oder wir malen sie an die Tafel.

Name : _____ Vorname : _____

Schule : _____ Klasse : _____

1. Zeichen verbinden

Hier siehst Du 2 Kreise und eine gerade Linie.



Deine Aufgabe ist es, aus diesen Zeichen möglichst viele verschiedene Figuren zu zeichnen.
Hier ist ein Beispiel:

Beachte, daß die Zeichen immer zusammenhängen müssen. Zeichne nun möglichst viele und möglichst verschiedene Figuren.

Jetzt machst Du es mit diesen 3 Zeichen genauso.

Bitte nicht umblättern!

Überschriften

Du bekommst jetzt einige Überschriften, so wie sie über kleinen Geschichten stehen könnten. Schreibe alles auf, was Dir zu jeder Überschrift einfällt. Du sollst jedoch keine Geschichten schreiben, sondern nur möglichst viele Wörter, die Dir zu der Überschrift einfallen.

2. ' Kletterpartie '

3. ' In der Pause '

Bitte nicht umblättern!

4. Behausungen

Zeichne möglichst viele verschiedene Arten von Behausungen, in denen Menschen wohnen können. Die Menschen müssen in diesen Behausungen Schutz vor Kälte, Regen und Schnee finden und sie müssen darin kochen, essen und schlafen können.
Es kommt nicht darauf an, Einzelheiten und Ausschmückungen zu zeichnen, sondern darauf, daß Du Dir möglichst viele Arten von Behausungen einfallen läßt. Damit wir auch die ausgefallensten Einfälle erkennen können, schreibe bitte neben Deine Zeichnungen, was sie darstellen sollen.

Bitte nicht umblättern !

5. Rechenaufgaben

Bilde aus den untenstehenden Zahlen möglichst viele verschiedene Rechenaufgaben. Beachte, daß in jeder Rechenaufgabe eine Zahl nur einmal vorkommen darf.

1, 2, 3, 4, 5, 6, 7, 8, 9, 0

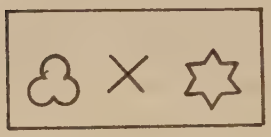
Beispiel: 2 + 3 = 5

Erfinde nun weitere Rechenaufgaben.

Bitte nicht umblättern!

6. Figurenfinden

In diesem Spiel sollst Du zu 3 Figuren ähnliche Figuren finden, die dazu passen. Schau Dir die folgenden Figuren an. Die erste Figur besteht aus kleinen Rundungen, die zweite besteht aus einfachen Strichen und die dritte Figur ist zackig. Zeichne nun unter jede Figur ähnliche, die dazu passen.



Bitte nicht umblättern!

Verwendungen

Jetzt sollst Du Dir viele Verwendungsmöglichkeiten für Gegenstände ausdenken. Denke z.B. an einen Bleistift! Was kann man mit einem Bleistift alles machen? Zunächst einmal schreiben. Das fällt natürlich jedem gleich ein. Aber was kann man noch alles mit einem Bleistift machen?
Man kann ihn als Zeigestab benutzen, man kann damit die Erde in einem Blumentopf auflockern oder ein Pfeifenraucher kann damit den verbrannten Tabak aus der Pfeife herausbefördern. Man kann also auch an ausgefallene Verwendungsmöglichkeiten denken, für die man normalerweise keinen Bleistift benutzt.

7. Nun schreibe einmal in kurzen Worten auf, was man alles mit einem Gürtel machen kann.

Bitte nicht umblättern!

8. Und jetzt schreibe einmal auf, was man mit einem Kleiderbügel alles machen kann.

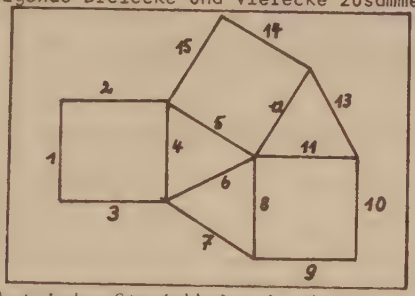
---

9. Zum Schluß Überlege Dir noch, was man mit einer Zeitung alles machen kann.

Bitte nicht umblättern!

10. Streichhölzer

Jetzt kommt wieder etwas ganz anderes. Stell Dir vor, Du hättest mit Streichhölzern folgende Dreiecke und Vierecke zusammengesetzt.



Wie Du sehen kannst, hat jedes Streichholz eine bestimmte Nummer. Deine Aufgabe ist es nun, in Gedanken soviele Streichhölzer wegzunehmen, daß immer 2 Quadrate und 1 Dreieck übrigbleiben. Es gibt natürlich viele verschiedene Möglichkeiten, dies zu tun. Laß Dir möglichst viele einfallen und schreibe dabei immer die Nummern der Streichhölzer in eine Reihe, die wegfallen. Es kommt also darauf an, daß Du möglichst viele Möglichkeiten findest, die Streichhölzer wegzunehmen.

Bitte nicht umblättern!

Geschenke zaubern

Stell Dir vor, Du könntest die herrlichsten und ausgefallensten Geburtstagsgeschenke zaubern, Dinge, die außer Dir vielleicht gar niemand verschenken kann.
Welche drei Geschenke würdest Du zaubern?

11. Zum Geburtstag Deines Freundes oder Deiner Freundin

---

12. Zum Geburtstag Deiner Mutter

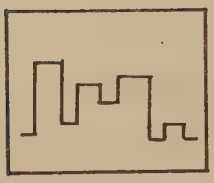
Bitte nicht umblättern!

13. Bilderraten

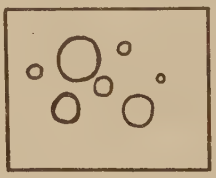
Nun folgen einige Bilder, die nicht fertig gezeichnet wurden. Schau Dir die Bilder genau an und schreibe dann darunter, was das sein könnte, wenn die Zeichnung fertig wäre. Laß Deiner Phantasie ruhig freien Lauf und schreibe möglichst viele Dinge auf.

14.

15.



16.



Bitte nicht umblättern !

Vergleiche

Auf dieser Seite stehen verschiedene lustige Vergleiche zwischen zwei Gegenständen oder zwei Tätigkeiten.

Beispiel: Schwimmen ist wie radfahren, weil .....

weil man für beides Arme und Beine braucht
weil beides Spaß macht
weil man sich durch beides fortbewegen kann
weil beides Sportarten sind usw.

Wie Du siehst, kommt es darauf an, möglichst viele Gründe für die Vergleiche zu finden.

Nun überlege Dir, warum man ein Nest mit einem Bett vergleichen kann.

17. Ein Nest ist wie ein Bett, weil .....

---

18. Sprechen ist wie Singen, weil .....

---

19. Eine Kerze ist wie ein Ofen, weil .....

Bitte nicht umblättern!

20. Buchstabenpuzzle

Du findest nun eine Reihe von Buchstaben, aus denen Du Gegenstände zeichnen sollst, die etwas bedeuten. Du darfst die gleichen Buchstaben mehrmals verwenden, aber keine Linien und Striche hinzufügen.

C E H I O P T U X Y

Beispiel:

Dieser Gegenstand entsteht, wenn man ein U und ein T zusammenfügt. Versuche nun, weitere Gegenstände zu finden, und zeichne sie auf diese Seite.

Bitte nicht umblättern!

21. Streit schlichten

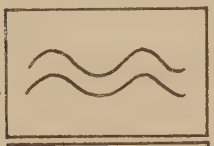
Volker ist in einer schwierigen Lage. Zwei größere Jungen, die er beide gut leiden kann, verprügeln sich, und er möchte die Schlägerei so schnell wie möglich beenden. Er ist schwächer als die beiden und kann daher die Kampfhähne nicht mit Gewalt trennen. Was könnte er sonst tun, um den Streit zu beenden? Schreibe in kurzen Worten alle Möglichkeiten auf, die Dir einfallen!

Bitte nicht umblättern!

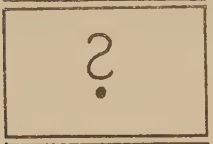
22. Berufe raten

Für dieses Spiel haben wir uns 5 Zeichen ausgedacht, die für eine Reihe von Berufen stehen könnten. Du sollst nun zu jedem einzelnen Zeichen aufschreiben, welche Berufe damit gemeint sein könnten.

Es gibt mehrere Möglichkeiten.

1. 

2.

3. 

4.

5.

1. _____
_____
_____

2. _____
_____
_____

3. _____
_____
_____

4. _____
_____
_____

5. _____
_____
_____

Dritte Versuchsform (IV,1)

Erziehungswissenschaftliche Hochschule Rheinland-Pfalz
Zentrum für empirische pädagogische Forschung

Versuchsform IV,1

Du hast sicher schon an einem oder sogar an mehreren Tests teilgenommen. Erinnerst Du Dich noch daran? Vermutlich hast Du bei einigen der Aufgaben ein bißchen Angst gehabt, daß Du sie nicht lösen kannst. Auf jeden Fall war es fast immer so, daß Deine Lösungen entweder richtig oder falsch waren.
Dieses Mal ist das anders, es gibt nämlich gar keine falschen Antworten. Du kannst die Aufgaben auf diesen Blättern höchstens mehr oder weniger schlau lösen. Besonders gut kannst Du sie lösen, wenn Du Deiner Phantasie freien Lauf läßt, Dir möglichst viele Gedanken einfallen läßt und auch ausgefallene Antworten gibst, an die vielleicht außer Dir niemand denkt. Schau also nicht nach, was Dein Nachbar schreibt oder zeichnet, sondern denk Dir selbst etwas aus! Übrigens, die Schrift und die Rechtschreibung sind heute nicht wichtig. Die Hauptsache ist, daß wir hinterher wissen, was Du gemeint hast.
Du wirst sehen, daß recht lustige Aufgaben dabei sind, die Dir sicher Spaß machen werden. Wahrscheinlich findest Du sogar, daß das gar keine richtigen Aufgaben, sondern eher Spiele sind.

Name: _____ Vorname: _____

Schule: _____ Klasse: _____

1. Spielklötze

In unserem ersten Spiel stellst Du Dir drei Bauklötze vor, so wie sie hier gezeichnet sind.

Mit diesen Klötzen baust Du in Gedanken alle Figuren, die Dir einfallen. Zeichne dann die Figuren auf diese Seite. (Die Linien auf dem Papier helfen Dir, daß Du die Figuren besser zeichnen kannst; ein Lineal brauchst Du dabei nicht.)

Wenn man baut, soll nichts umfallen. Zeichne Deine Figuren auch so, daß die Klötze stehen bleiben.

Versuche nun, möglichst viele verschiedene Figuren zu zeichnen.

Bitte nicht umblättern!

Verrückte Geschichten

Jetzt wollen wir uns einmal ganz verrückte Fragen überlegen. Stell Dir einmal vor, es würde keinen Winter mehr geben. Was würde dann alles geschehen? Man könnte z.B. nicht mehr Schlitten fahren, die Leute könnten ihre Winterkleider wegwerfen, der Kohlenhändler müßte sich eine andere Arbeit suchen usw.

2. Nun stell Dir einmal vor, alle Ameisen wären zwei Meter groß. Was könnte sich daraus alles ergeben?

3. Nimm einmal an, Du könntest nicht nur stehen, gehen, sitzen und liegen, sondern Du könntest auch frei in der Luft schweben. Was könnte da alles geschehen?

Bitte nicht umblättern!

4. Behausungen

Zeichne möglichst viele verschiedene Arten von Behausungen, in denen Menschen wohnen können. Die Menschen müssen in diesen Behausungen Schutz vor Kälte, Regen und Schnee finden und sie müssen darin kochen, essen und schlafen können.
Es kommt nicht darauf an, Einzelheiten und Ausschmückungen zu zeichnen, sondern darauf, daß Du Dir möglichst viele Arten von Behausungen einfallen läßt. Damit wir auch die ausgefallensten Einfälle erkennen können, schreibe bitte neben Deine Zeichnungen, was sie darstellen sollen.

Bitte nicht umblättern!

Lustige Vergleiche

Auf dieser Seite stehen lustige Vergleiche zwischen zwei Gegenständen.
Beispiel: Ein Auto ist wie ein Fahrrad, weil.....
weil beide Räder haben,
weil man mit beiden schneller vom Fleck kommt,
weil beide auf Straßen fahren,
weil Autofahren und Radfahren vielen Leuten Spaß macht usw.

Wie Du siehst, kommt es darauf an, möglichst viele Gründe für die Vergleiche zu finden.

Nun überlege Dir, warum man ein Nest mit einem Bett vergleichen kann.

5. Ein Nest ist wie ein Bett, weil .....

---

6. Eine Kerze ist wie ein Ofen, weil .....

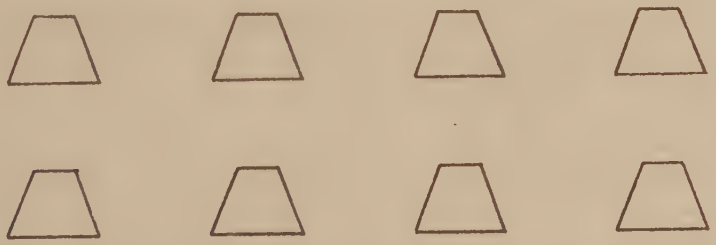
Bitte nicht umblättern!

Zeichnungen

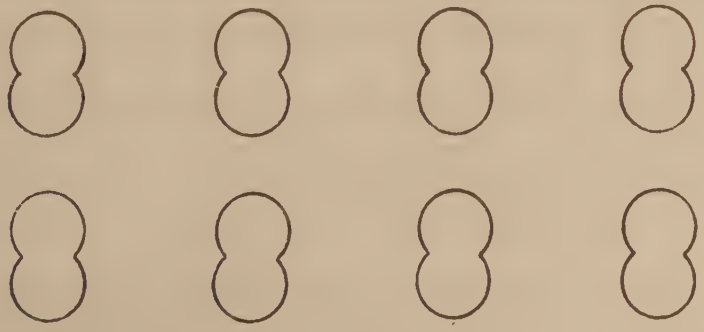
Nun kommen wir zum nächsten Spiel. Du siehst hier immer die gleichen Figuren. Zeichne darauf bekannte Dinge, indem Du andere Zeichen oder Striche hinzufügst. Es kommt jedoch nicht darauf an, daß Du sehr schön zeichnest, sondern es ist wichtig, daß Dir möglichst viele verschiedene Dinge einfallen. Damit wir erkennen können, was Du gemeint hast, schreibe zu jeder Zeichnung dazu, was es sein soll.

7.

8.



9.



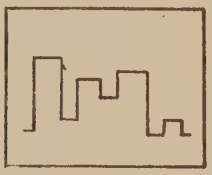
Bitte nicht umblättern!

Bilderraten

Zum Schluß haben wir noch zwei Bilder, die nicht fertiggezeichnet wurden. Schau Dir die Bilder genau an und schreibe dann darunter, was das sein könnte, wenn die Zeichnung fertig wäre. Laß Deiner Phantasie ruhig freien Lauf und schreibe möglichst viele Dinge auf!

10.

11.



Anleitung zur Durchführung der Versuchsform IV,1

1. Vorbereitung der Untersuchung

Der Versuchsleiter (Vl) sorgt dafür, daß die Testdurchführung in einem ruhigen Raum stattfindet und keine Störungen von außen erfolgen. Es ist ratsam, an der Tür ein Schild "Bitte nicht stören! Testdurchführung!" zu befestigen.

Jedes Kind braucht einen Bleistift. Der Vl sollte einige Reservebleistifte und einen Bleistiftspitzer bereithalten.

2. Durchführung der Untersuchung

Die Versuchsform IV,1 ist eine Zusammenstellung von Aufgaben, die die Fähigkeit zu divergentem Denken als einer von mehreren Vorbedingungen für kreatives Verhalten erfassen sollen.

Entsprechend dieser Zielsetzung sollte die Atmosphäre während der Untersuchung möglichst entspannt und gelöst sein. Die leider unvermeidliche Zeitbegrenzung sollte eingehalten werden, ohne daß die Kinder viel davon merken, also ohne Stoppuhr (eine Armbanduhr mit Sekundenzeiger wäre unbedingt vorzuziehen) und ohne allzu strenges Abbrechen der gerade ablaufenden Aktivität.(statt der bei Tests üblichen Formulierungen wie z.B."Halt! Legt die Bleistifte hin! Ich erkläre Euch jetzt die nächste Aufgabe." sollte der VL beispielsweise sagen:"So, nun habt Ihr Euch mit diesem Spiel lange genug beschäftigt, jetzt kommt ein anderes an die Reihe! Legt doch bitte Eure Bleistifte hin, ich möchte Euch nun erklären, wie das nächste Spiel geht".)
Daß die Kinder abschreiben, sollte natürlich nach Möglichkeit verhindert werden, aber nicht durch strenge Ermahnungen, sondern durch den freundlichen Hinweis, daß das keine Klassenarbeit sei und daß man doch nicht abzuschreiben brauche, wenn es gar keine falschen Antworten gibt.

Wenn die Kinder bei der einen oder anderen Aufgabe lachen und lebhaft werden, sollte der Vl lieber einige wenige Minuten opfern und sich an der

allgemeinen Heiterkeit beteiligen, als im Interesse einer zügigen Untersuchungsdurchführung jegliches spontane Verhalten zu unterdrücken. Wenn die Kinder allerdings mögliche Aufgabenbeantwortungen nennen, soll er sie in möglichst spielerischer Form dazu bringen, ihre Lösungen für sich zu behalten und aufzuschreiben (z.B. :"Psst! Haltet Eure Einfälle streng geheim! Große Erfinder erzählen auch nicht gleich allen Leuten, woran sie gerade arbeiten." o.ä.)

Der Vl verteilt die Testhefte und bittet die Kinder, sie noch nicht aufzuschlagen. Wenn alle Kinder ihre Testhefte erhalten haben, gibt der Vl eine kurze Einführung, wobei er den Text auf der Titelseite des Aufgabenhefts übernehmen kann, wenn er statt der Anrede in der 2.Person Singular die 2.Person Plural verwendet (also: "Ihr habt sicher schon...."). Die Kinder können zuhören oder mitlesen.
Der Text sollte möglichst wörtlich übernommen werden; inhaltliche Auslassungen oder Hinzufügungen würden die Untersuchungsbedingungen von Klasse zu Klasse verschieden gestalten, was unbedingt vermieden werden muß.
Falls der Vl es deshalb vorziehen sollte, den Text vorzulesen, sollte das in möglichst gelockerter und lebhafter Weise geschehen.

Die Kinder sollen dann zunächst Name, Vorname, Schule und Klasse auf die Titelseite schreiben, wobei der Vl Schule und Klasse angibt, damit sie einheitlich aufgeschrieben werden. Die Anweisungen zu den einzelnen Aufgaben werden vom Vl mündlich gegeben, wobei er - bei veränderter Anrede - die Formulierungen aus dem Aufgabenheft übernimmt. Das für den Vl bestimmte Aufgabenheft - handschriftlich als "Testleiterexemplar" gekennzeichnet - enthält einige zusätzliche Angaben und Hinweise für die Schüler. Außerdem ist darin die Zeitbegrenzung für jede einzelne Aufgabe angegeben.

Die Durchführung der Untersuchung dürfte etwa 45 Minuten dauern.

Für die Mitarbeit der Versuchsleiter und der Schüler danken wir sehr herzlich.

Erste und zweite (gekürzte) Version des Lehrerfragebogens

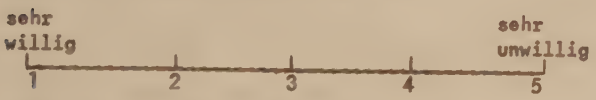
Fragebogen zum Unterrichtsverhalten der Schüler

Wir bitten Sie, das Verhalten Ihrer Schüler in bestimmten schulischen Situationen zu beurteilen.

Zur Erleichterung Ihrer Aufgabe werden Ihnen Skalen vorgegeben, auf denen Sie Ihre Beurteilung durch Ankreuzen angeben können. Teilen Sie uns also bitte das Ergebnis Ihrer Beurteilung dadurch mit, daß Sie eine der Zahlen zwischen den Wortpaaren ankreuzen.

Zur Verdeutlichung des Vorgehens ein Beispiel:

Arbeitsaufträge befolgt der Schüler



Kommt ein Schüler Ihren Arbeitsaufträgen im allgemeinen mit großem Eifer nach, würden Sie die Zahl 1 ankreuzen. Befolgt ein Schüler Ihre Anweisungen ausgesprochen ungern und nur unter Anwendung von Druck, würden Sie die 5 ankreuzen. Trifft weder das eine noch das andere zu, wäre die mittlere Zahl 3 anzukreuzen, die für einen allgemeinen Durchschnitt der Schüler dieser Klassenstufe kennzeichnend ist. Durch die Zahl 2 bzw. 4 wird die eher willige bzw. eher unwillige Befolgung von Aufträgen angezeigt, die Zahlen 2 und 4 symbolisieren also schwächere Ausprägungsgrade von 1 und 5.

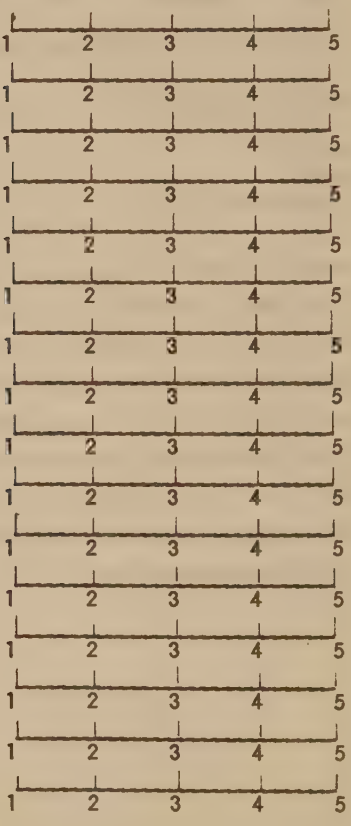
Zur ersten Frage wird ein Teil der Namen und Skalen für eine Klasse als Beispiel aufgeführt, die restlichen Fragen werden ohne die zugehörigen Namen und Skalen mitgeteilt.

1. An Tätigkeiten, bei denen die saubere und sorgfältige Ausführung im Vordergrund steht, ist der Schüler

Beurteilungsskala:

Name:

sehr interessiert — völlig uninteressiert



2. Bei neuen oder ungewöhnlichen Fragestellungen meldet er sich
sehr häufig sehr selten oder nie

3. Diszipliniertes, angepaßtes Verhalten fällt dem Schüler
sehr schwer sehr leicht

4. Seine Vorliebe für Aufgaben, deren Lösungsweg von den Schülern relativ selbständig erarbeitet werden soll, ist
sehr groß sehr gering

5. Der Schüler übernimmt Meinungen und Ansichten von Klassenkameraden
sehr häufig sehr selten oder nie

6. An Unterrichtsgesprächen, in denen Einfallsreichtum und originelle Ideen eine große Rolle spielen, beteiligt er sich
sehr stark sehr schwach

7. Bei neuen Arbeitsaufträgen bittet der Schüler um zusätzliche Hilfen und Erklärungen
sehr häufig sehr selten oder nie

8. Aufgaben zur routinemäßigen Einübung einer neu erworbenen Fertigkeit macht er sehr gern sehr ungern

9. Bei der Erarbeitung eines Unterrichtsthemas bringt der Schüler eigene Vorschläge und Anregungen
sehr häufig sehr selten

10. Er stört den Unterricht durch Fragen, Bemerkungen und Einwände, die vom Thema wegführen,
sehr häufig sehr selten oder nie

11. Bei der Verfolgung selbstgesteckter Ziele ist der Schüler
sehr ausdauernd nicht ausdauernd

12. Bei der Darstellung eines Sachverhalts (durch den Lehrer, durch Mitschüler oder Bücher) entdeckt der Schüler offene Fragen oder Probleme sehr häufig sehr selten oder nie

13. Der Schüler stellt sich selbst Aufgaben oder steckt sich eigene Ziele sehr häufig sehr selten oder nie

1. Bitte stufen Sie das Interesse Ihrer Schüler an Tätigkeiten ein, bei denen die saubere und sorgfältige Ausführung im Vordergrund steht.

2. Bitte beurteilen Sie, wie häufig sich Ihre Schüler bei neuen oder ungewöhnlichen Fragestellungen melden.

3. Beurteilen Sie bitte, ob Ihren Schülern diszipliniertes, angepaßtes Verhalten leicht oder schwer fällt.

4. Bitte stufen Sie die Vorliebe Ihrer Schüler für Aufgaben ein, deren Lösungsweg relativ selbständig erarbeitet werden muß.

5. Bitte stufen Sie die Beteiligung Ihrer Schüler an Unterrichtsgesprächen ein, in denen Einfallsreichtum und originelle Ideen eine große Rolle spielen.

6. Beurteilen Sie bitte, wie gerne Ihre Schüler Aufgaben zur routinemäßigen Einübung einer neu erworbenen Fertigkeit machen.

7. Bitte überlegen Sie, wie häufig sich Ihre Schüler selbst Aufgaben stellen oder wie häufig sie sich eigene Ziele setzen.

8. Stufen Sie bitte die Häufigkeit ein, mit der Ihre Schüler bei der Erarbeitung eines Unterrichtsthemas eigene Vorschläge und Anregungen bringen.

9. Beurteilen Sie bitte, ob Ihre Schüler bei der Darstellung eines Sachverhalts (durch den Lehrer, durch Mitschüler oder Bücher) offene Fragen oder Probleme entdecken.

Die Vorgabe der Namen und Skalen erfolgte in der gleichen Form wie bei der ersten Version des Fragebogens.

Materialien zur Schülerbefragung

Fragebogen für Schüler

(Anweisung für den Lehrer)

1.Teil

Stellt Euch bitte einmal vor, wir würden gemeinsam einige Tage ins Landschulheim fahren!
Was wir tagsüber unternehmen, würde ich mir überlegen, aber für Unterhaltung am Abend müßtet Ihr schon selbst sorgen. Schön wäre es natürlich, wenn wir für jeden Abend ein anderes Programm hätten, das einige von Euch möglichst selbständig - also ohne meine Hilfe - ausarbeiten könnten. Vielleicht könntet Ihr an einem Abend basteln, und an einem anderen Theater spielen ?

Damit wir aber immer diejenigen von Euch mit der Gestaltung des Abends beauftragen, die für das jeweilige Programm auch wirklich die tollsten Einfälle haben und die besten Vorschläge machen können, habe ich mir ein paar Fragen überlegt, die Ihr mir beantworten sollt.
Denkt bitte daran, daß ich von Euch nicht wissen will, wer in der Schule am besten ist; das weiß ich ja selbst und es ist gar nicht gesagt, daß gerade die besten Schüler auch die besten Einfälle zu Eurer Unterhaltung hätten. Und noch eins : da wir ganz unterschiedliche Programme haben, sind es vielleicht auch verschiedene Gruppen, die an den Abenden die Programmgestaltung übernehmen sollten.

Zunächst suchen wir Schüler, die Euch Vorschläge machen sollen, was Ihr an dem Bastelabend basteln könntet.
Ihr bekommt jetzt gleich Namenskärtchen von allen Schülern dieser Klasse, Euer eigener Name steht auch auf einem Kärtchen. Und außerdem bekommt jeder von Euch ein Blatt Papier, das so gefaltet und geheftet ist, daß 5 kleine Taschen entstanden sind. Auf dem oberen Täschchen steht :
Hierher gehören alle Schüler, die das ganz besonders gut können.
Auf dem zweiten Täschchen steht :
Hier hinein kommen alle Schüler, die das auch noch recht gut können.

Auf dem mittleren Täschchen steht :
Hier ist der Platz für die Schüler, die das weder gut noch schlecht, sondern so mittelmäßig können.

Auf dem vierten Täschchen steht :
Hierher gehören diejenigen Schüler, die das eher schlecht können.

Auf dem unteren Täschchen steht :
Hier hinein kommen alle Schüler, die so etwas überhaupt nicht können.

So, und nun überlegt einmal, in welche Täschchen Ihr die Namen Eurer Mitschüler und Euren eigenen stecken müßt, wenn die Frage lautet :
Wer kann viele interessante Vorschläge machen, was Ihr an dem Bastelabend basteln könntet ?

Denkt bitte daran, daß nicht nach geschickten Bastlern gefragt wird, sondern nach Schülern, denen viele tolle Dinge einfallen, die Ihr basteln könntet.

Schreibt bitte rechts oben (zeigen!) Euren Vor- und Zunamen auf das Blatt, das ich jetzt austeile !

Bitte die Blätter beim Einsammeln quer halten, damit die Namenskärtchen nicht herausrutschen !

2.Teil :

Ihr erinnert Euch sicher an unser Gespräch über eine Fahrt ins Landschulheim, bei der Ihr - wenn eine solche Fahrt zustande käme - Euren Zeitvertreib am Abend selbst in die Hand nehmen würdet. Ich sagte schon, daß Ihr an einem Abend vielleicht Theater spielen könntet. Natürlich würdet Ihr Euch das Theaterstück, die Personen und die Handlung, also was da alles passiert, selbst ausdenken.
Wie wäre es z.B., wenn Ihr Euch einige lustige Begebenheiten ausdenken würdet, die passieren, wenn Herr und Frau Schmittchen mit ihren 4 Kindern, der Oma und einem Kater im nagelneuen Kleinwagen in die Ferien fahren, und zwar nach Italien, obwohl kein Mitglied der Familie italienisch spricht ?
Über Planung und Verlauf dieser Ferienreise könnte man doch sicher eine Komödie, also ein lustiges Theaterstück, machen !
Nun kommt meine Frage dazu :
Wer von Euch hätte wohl beim Ausdenken eines Theaterstückes die besten Ideen ?

Ihr bekommt wieder die Blätter mit den kleinen Taschen und die Namenskärtchen aller Schüler dieser Klasse.
Eure Aufgabe für heute ist es, Euch zu überlegen, wer von Euch beim Ausdenken eines spannenden Theaterstückes ganz besonders gut ist, wer das auch noch recht gut machen könnte, wer da so mittelmäßig ist, wer das eher schlecht machen würde und wer das überhaupt nicht könnte.

Also, denkt daran, wir suchen jetzt keine guten Schauspieler, sondern Schüler, die sich die Handlung Eures Theaterstückes ausdenken !
Vergeßt nicht bitte, Eure Namen in die rechte obere Ecke zu schreiben !

3. Teil :

Zum Schluß hätte ich noch eine ganz andere Frage :

Ihr wißt doch, daß die Schüler im Landschulheim in Schlafsälen schlafen. Und vielleicht könnt Ihr Euch denken, daß da am Abend viele lustige Streiche ausgeheckt werden ! Da werden Schlafanzüge versteckt und Bürsten unter das Leintuch gelegt und vieles andere wird angestellt !
Manche Streiche sind schon so bekannt, daß viele Schüler gar nicht mehr darauf hereinfallen ; sie tasten Kopfkissen und Leintuch nach stacheligen Gegenständen ab und sehen sich ihren Schlafanzug daraufhin an, ob die Ärmel oder Hosenbeine zugenäht oder verknotet sind.
Aber es gibt immer wieder Schüler, denen etwas ganz Neues einfällt und gegen diese neu ausgedachten Streiche kann man sich dann nicht vorsehen, sondern man fällt darauf herein und es gibt meist ein großes Gelächter.

Nun sollt Ihr mir verraten, wer von Euch im Ausdenken neuer und lustiger Streiche vermutlich ganz besonders gut ist, wer das auch noch recht gut kann, wer da so mittelmäßig ist usw.
Es wird genau so gemacht wie bei den anderen Fragen, nur geht es dieses Mal um das Streiche-Ausdenken.

Schreibt bitte wieder Eure Namen rechts oben auf das Blatt !

Wenn die Schüler mit dem Einordnen der Namenskärtchen fertig sind, sagt der Lehrer: "Bevor ich jetzt die Blätter einsammele, schreibt mir bitte noch rechts unten (bitte zeigen!) die Namen Eurer 3 besten Freunde oder Freundinnen auf das Blatt !"

Name:

Wer kann viele interessante Vorschläge machen, was Ihr an einem Bastelabend basteln könntet?

Hierher gehören alle Schüler, die das ganz besonders gut können.

Hier hinein kommen alle Schüler, die das auch noch recht gut können.

Hier ist der Platz für die Schüler, die das weder gut noch schlecht, sondern so mittelmäßig können.

Hierher gehören diejenigen Schüler, die das eher schlecht können.

Hier hinein kommen alle Schüler, die so etwas überhaupt nicht können.

Name:

Hierher gehören alle Schüler, die das ganz besonders gut können.

Hier hinein kommen alle Schüler, die das auch noch recht gut können.

Hier ist der Platz für die Schüler, die das weder gut noch schlecht, sondern so mittelmäßig können.

Hierher gehören diejenigen Schüler, die das eher schlecht können.

Hier hinein kommen alle Schüler, die so etwas überhaupt nicht können.

Wer könnte sich eine spannende Handlung für ein Theaterstück ausdenken?

Name:

Hierher gehören alle Schüler, die das ganz besonders gut können.

Hier hinein kommen alle Schüler, die das auch noch recht gut können.

Hier ist der Platz für die Schüler, die das weder gut noch schlecht, sondern so mittelmäßig können.

Hierher gehören diejenigen Schüler, die das eher schlecht können.

Hier hinein kommen alle Schüler, die so etwas überhaupt nicht können.

Wer kann sich lustige Überraschungen und Streiche ausdenken?

Der netteste Schüler bzw. die netteste Schülerin in meiner Klasse ist:

Kreativitätstest von Wallach und Kogan

ERZIEHUNGSWISSENSCHAFTLICHE HOCHSCHULE RHEINLAND-PFALZ

Zentrum für empirische pädagogische Forschung

Versuchsform W und K

Zu Beginn des neuen Schuljahres bist Du sicher gut erholt und ausgeruht und hast noch nicht so viel zu tun, weil die Stundenpläne noch nicht fertig sind. Deshalb bitten wir Dich jetzt ein letztes Mal um Deine Mitarbeit: Du sollst eine Reihe lustiger Aufgaben bearbeiten, bei denen es wieder darum geht, daß Du Dir möglichst viele Antworten einfallen läßt.

Vor den großen Sommerferien hast Du schon einmal etwas Ähnliches gemacht, erinnerst Du Dich noch daran ? Auch dieses Mal gibt es keine falschen Antworten, sondern es kommt vor allem darauf an, daß Du Deiner Phantasie freien Lauf läßt, Dir möglichst viele Gedanken einfallen läßt und auch ausgefallene Antworten gibst, an die vielleicht außer Dir niemand denkt.

Name : _____ Vorname : _____

Schule : _____ Klasse : _____

Erstes Spiel "Dinge nennen"

In unserem ersten Spiel geht es darum, daß Du Dir möglichst viele Dinge mit gleichen Eigenschaften einfallen läßt und aufschreibst.

1) Schreibe alle runden Dinge auf, die Dir einfallen !

---

2) Schreibe alle Dinge auf, die Lärm und Krach machen !

---

3) Schreibe alle quadratischen Dinge auf, die Dir einfallen !

---

4) Schreibe alle Dinge auf, die Räder haben !

Zweites Spiel "Was bedeuten die Zeichnungen ?"

Versuche zu erraten, was die folgenden Zeichnungen darstellen sollen ! Laß Dir möglichst viel dazu einfallen und schreibe alles unter die dazugehörige Zeichnung. Wenn Dir zu einer Zeichnung gar nichts mehr einfällt, kannst Du bei der nächsten weitermachen.

Beispiel :

Was kann das sein ?

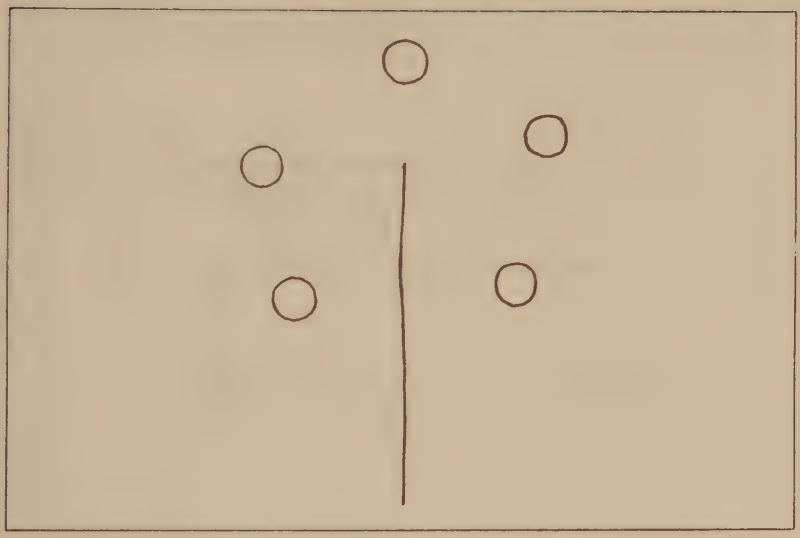
Stachelschwein, Augenbraue, Bürste, Igel, Sonnenuntergang, Nelke usw.

.....

.....

.....

3)



Was könnte das sein ?

.....

.....

.....

.....

.....

.....

.....

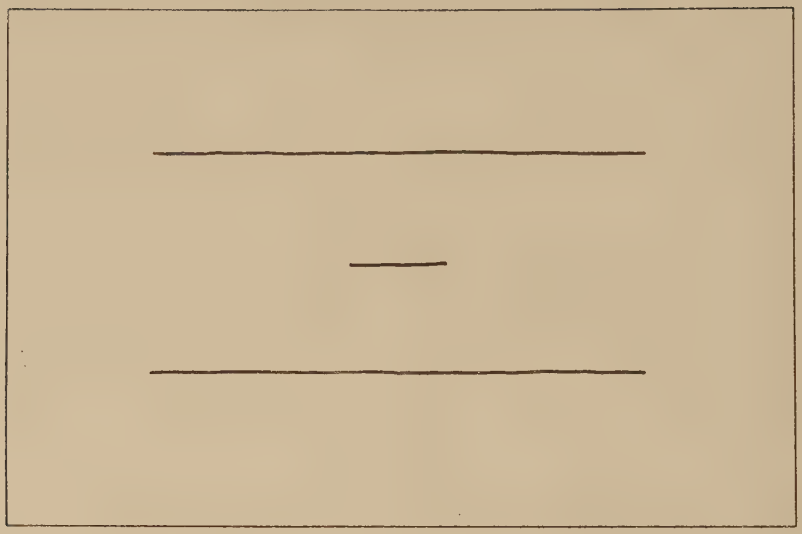
.....

.....

.....

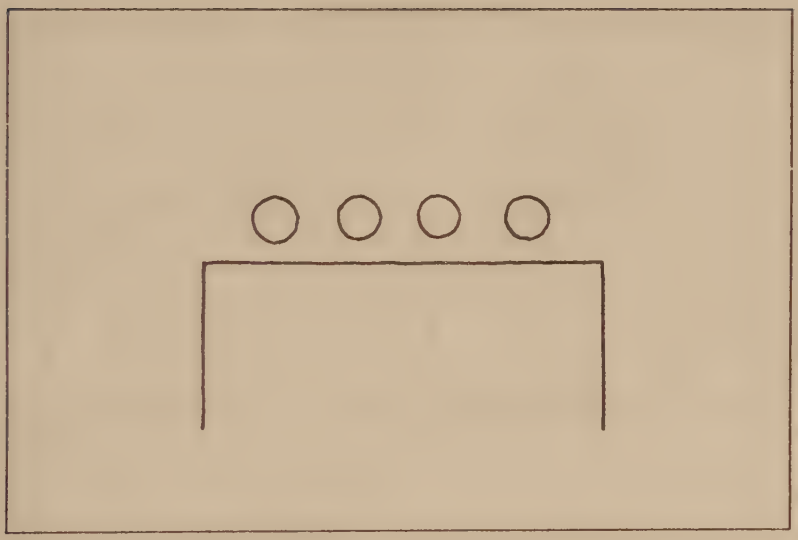
.....

6)



Was könnte das sein ?

7)



Was könnte das sein ?

.....

.....

.....

.....

.....

.....

.....

.....

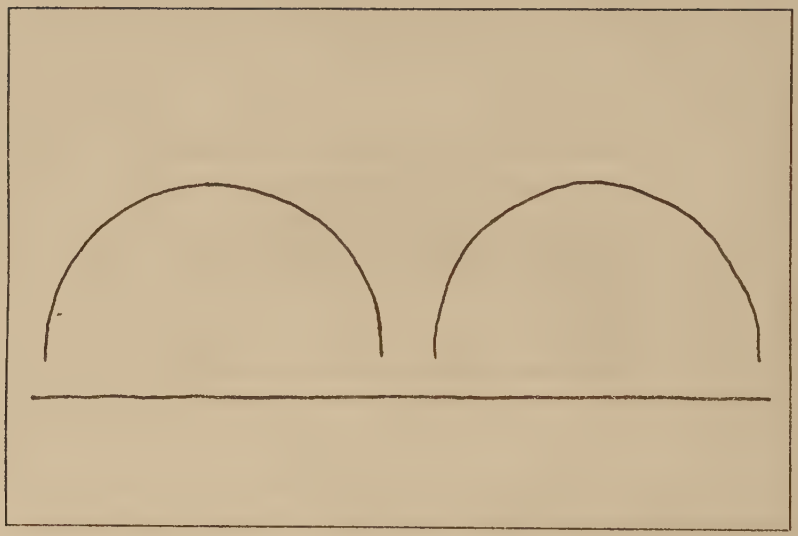
.....

.....

.....

.....

8)

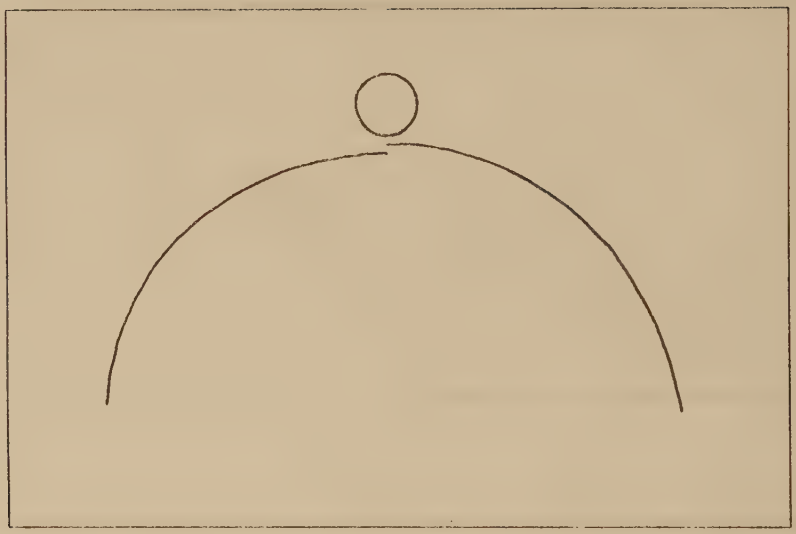


Was kann das sein ?

9)

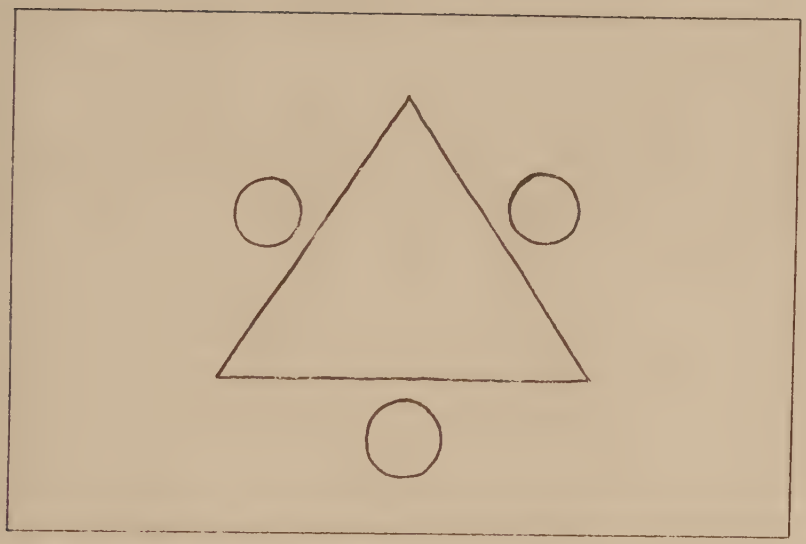
Was könnte das sein ?

10)



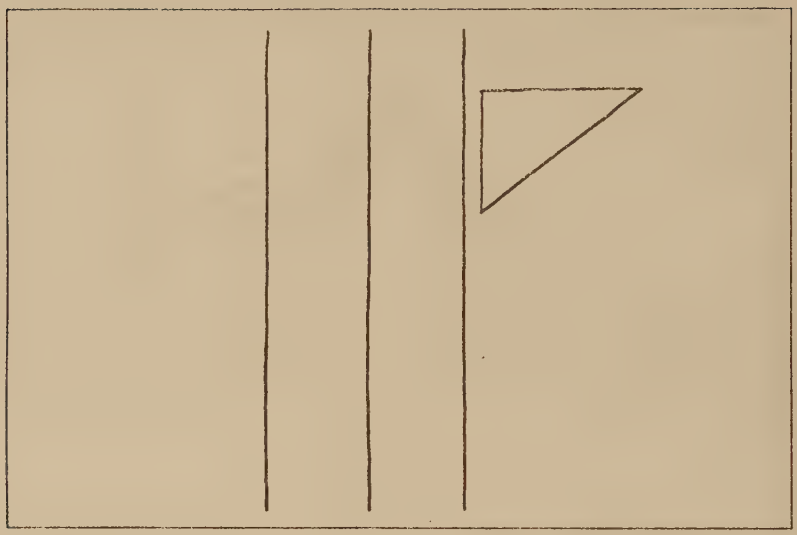
Was könnte das sein ?

11)



Was könnte das sein ?

12)



Was könnte das sein ?

Drittes Spiel "Verwendungen"

Bei diesem Spiel geht es darum, möglichst viele und möglichst verschiedene Verwendungsmöglichkeiten für bestimmte Gegenstände zu finden und aufzuschreiben.

13) Was kann man mit einer Zeitung alles anfangen ?

14) Was kann man mit einem Messer alles machen ?

15) Welche Verwendungsmöglichkeiten gibt es für einen Autoreifen ?

16) Was kann man alles mit einem Korken anfangen ?

17) Wozu kann man einen Schuh verwenden ?

18) Welche Verwendungsmöglichkeiten gibt es für einen Türschlüssel ?

19) Was kann man mit einem Hosenknopf alles machen ?

20) Wozu kann man einen Stuhl verwenden ?

Viertes Spiel "Was bedeuten die Linien ?"

Bei diesem Spiel kannst Du wieder Deiner Phantasie freien Lauf lassen. Deine Aufgabe ist es, zu erraten, was die folgenden 9 Abbildungen von verschiedenartigen Linien darstellen könnten. Schreibe alles auf, was Dir dazu einfällt !

21)

Was könnte das sein ?

.....

.....

.....

.....

.....

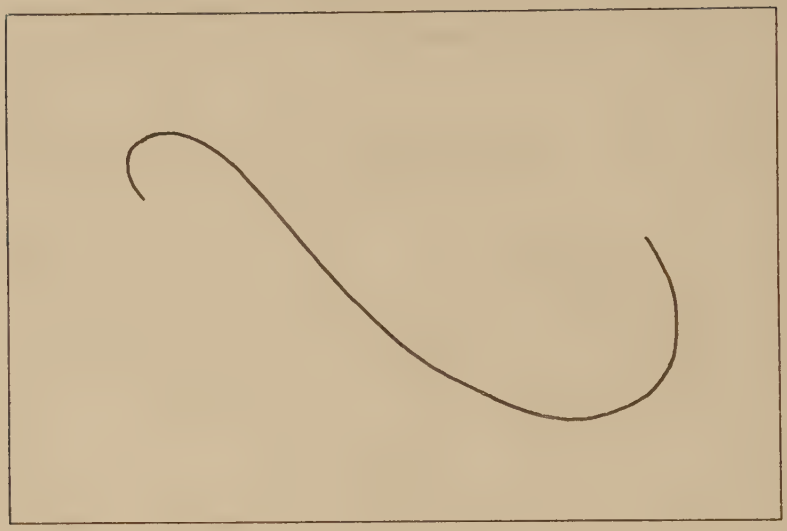
.....

.....

.....

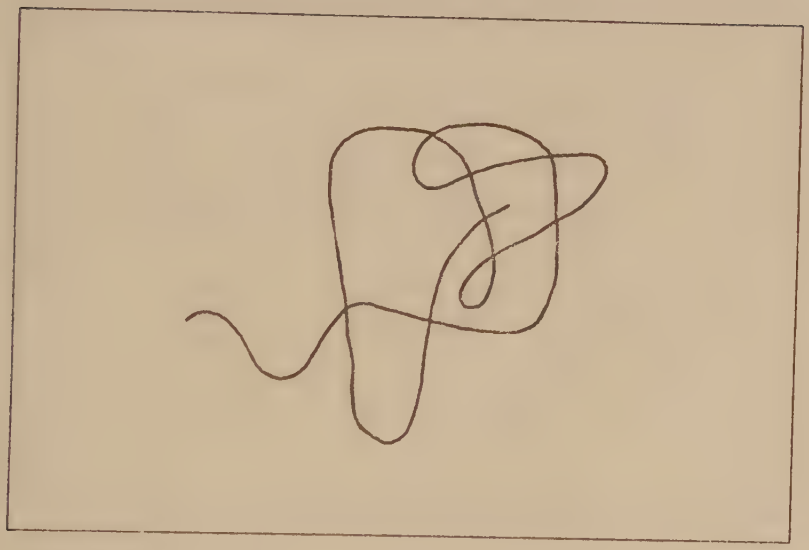
.....

22)



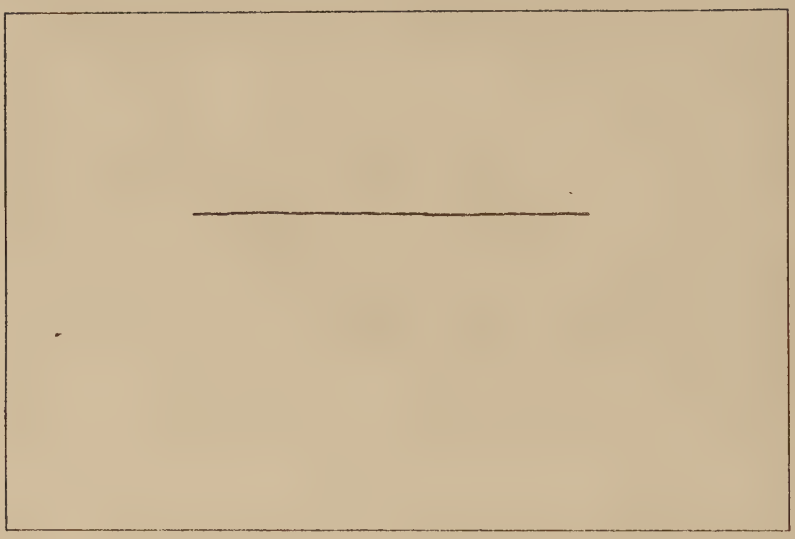
Was könnte das sein ?

23)



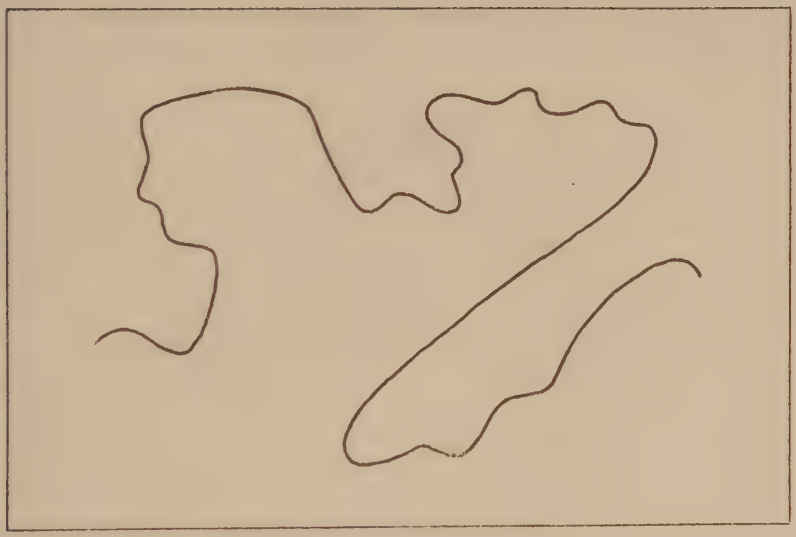
Was könnte das sein ?

24)



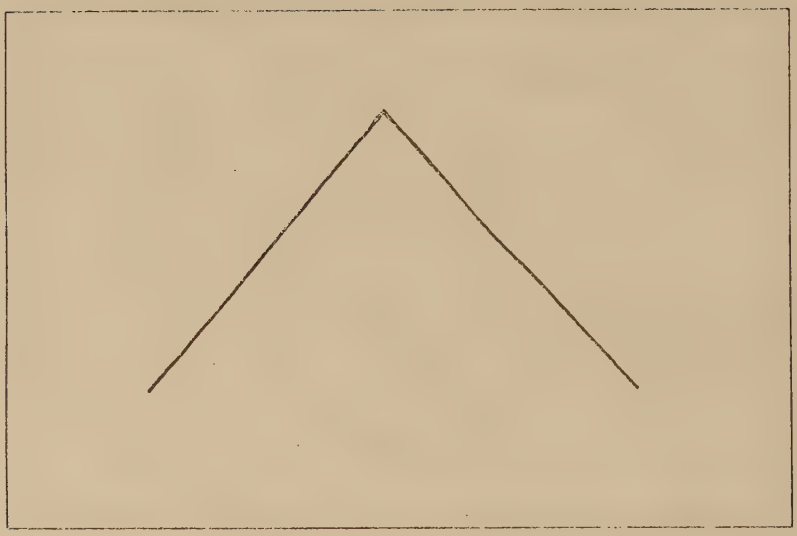
Was könnte das sein ?

25)



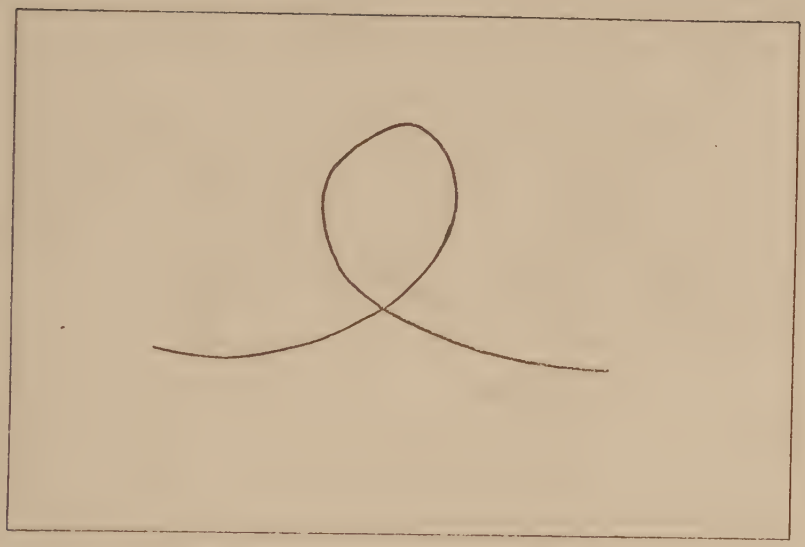
Was könnte das sein ?

16)



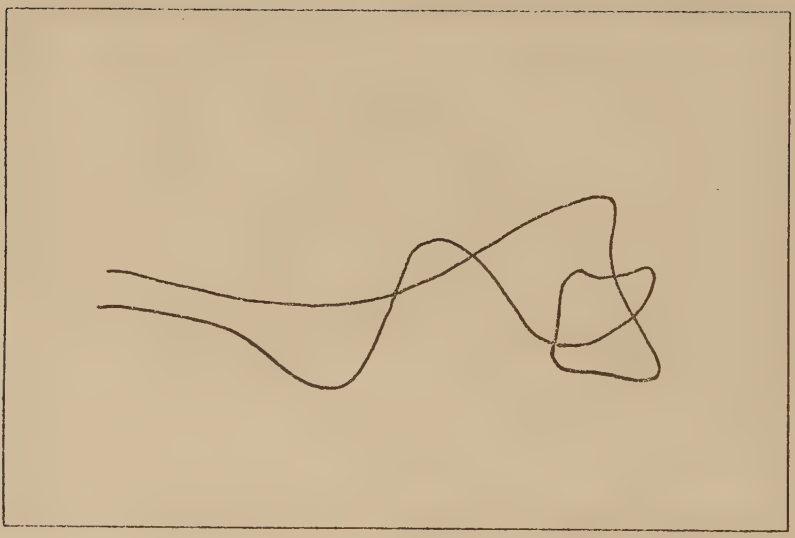
Was könnte das sein ?

27)



Was könnte das sein ?

28)



Was könnte das sein ?

.....

.....

.....

.....

.....

.....

.....

.....

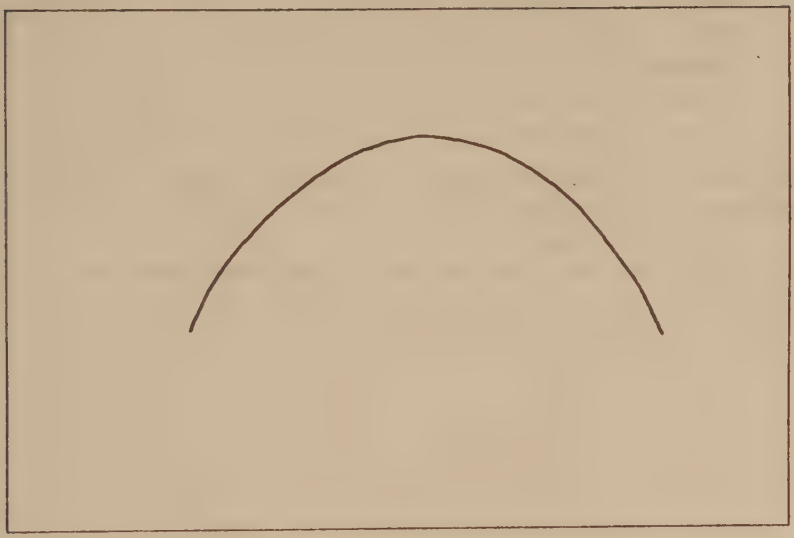
.....

.....

.....

.....

29)



Was könnte das sein ?

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

Fünftes Spiel " Gemeinsamkeiten suchen"

In unserem letzten Spiel geht es darum, herauszufinden, was an zwei Gegenständen alles gleich ist.

Beispiel : Was ist an einem Apfel und an einer Orange alles gleich ? Man kann beide essen, beide sind rund, beide sind Früchte, beide sollten saftig sein, aus beiden kann man Getränke machen, beide haben eine Schale, beide wachsen auf Bäumen, beide können faulen, beide kann man einen Abhang hinunterrollen lassen usw.

Schreibe bitte bei jeder der folgenden Fragen auf, was an den beiden Gegenständen alles gleich ist.

30) Was ist an einer Kartoffel und einer gelben Rübe (Karotte) alles gleich ?

---

31) Was ist an einer Katze und einer Maus alles gleich ?

---

32) Was ist an einer Eisenbahn und einem Traktor (Bulldog) alles gleich ?

---

33) Was ist an einem Lebensmittelgeschäft und einem Gasthaus alles gleich ?

---

34) Was ist an Milch und Fleisch alles gleich ?

---

35) Was ist an einer Geige und einem Klavier alles gleich ?

---

36) Was ist an einem Radio und einem Telefon alles gleich ?

---

37) Was ist an einer Uhr und einer Schreibmaschine alles gleich ?

---

38) Was ist an einem Vorhang und einem Teppich alles gleich ?

---

39) Was ist an einem Pult und einem Tisch alles gleich ?

Vierte Versuchsform (VI,K)

Erziehungswissenschaftliche Hochschule
Rheinland - Pfalz
Zentrum für empirische pädagogische Forschung

Versuchsform VI, K

Du hast sicher schon an einem oder sogar an mehreren Tests teilgenommen. Erinnerst Du Dich noch daran ? Vermutlich hast Du bei einigen Aufgaben ein bißchen Angst gehabt, daß Du sie nicht lösen kannst. Auf jeden Fall war es fast immer so, daß Deine Lösungen entweder richtig oder falsch waren.

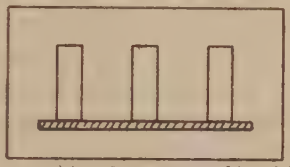
Dieses Mal ist das anders, es gibt nämlich gar keine falschen Antworten. Du kannst die Aufgaben auf diesen Blättern höchstens mehr oder weniger einfallsreich lösen. Besonders gut kannst Du sie lösen, wenn Du Deiner Phantasie freien Lauf läßt, Dir möglichst viele Gedanken einfallen läßt und auch ausgefallene Antworten gibst, an die vielleicht außer Dir niemand denkt. Schau also nicht nach, was Dein Nachbar schreibt oder zeichnet, sondern denk Dir selbst etwas aus !
Übrigens, die Schrift und die Rechtschreibung sind heute nicht wichtig. Die Hauptsache ist, daß wir hinterher wissen, was Du gemeint hast.
Du wirst sehen, daß recht lustige Aufgaben dabei sind, die Dir sicher Spaß machen werden. Wahrscheinlich findest Du sogar, daß das gar keine richtigen Aufgaben, sondern eher Spiele sind.

Name : _____ Vorname : _____

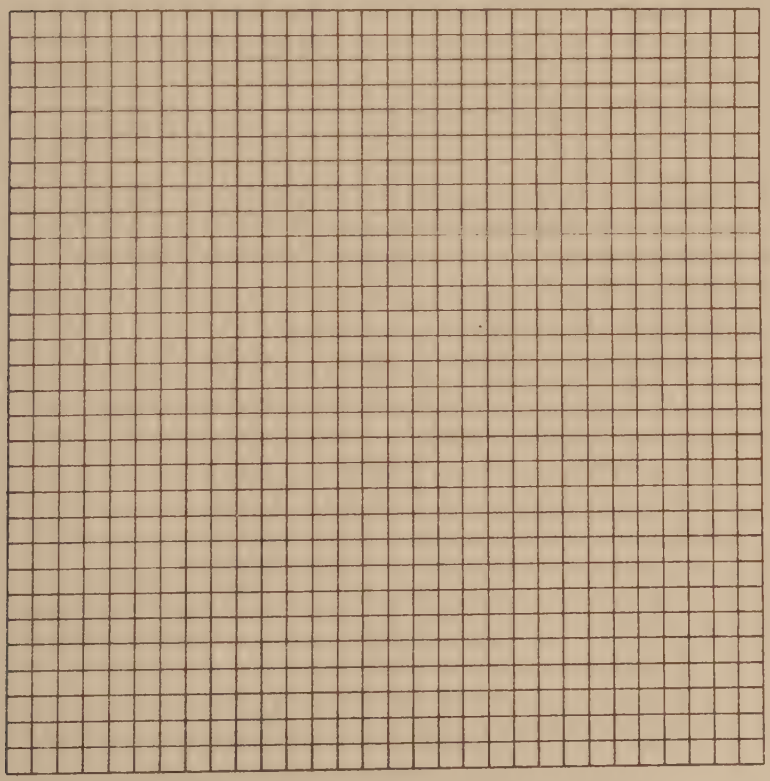
Schule: _____ Klasse : _____

1. Bauklötze

In unserem ersten Spiel stellst Du Dir drei Bauklötze vor, so wie sie hier gezeichnet sind. Du siehst, die Klötze sind länglich und stehen auf einem Brett.



Mit diesen Klötzen baust Du in Gedanken alle Figuren, die Dir einfallen. Zeichne dann die Figuren auf diese Seite. (Die Linien auf dem Papier helfen Dir, daß Du die Figuren besser zeichnen kannst; ein Lineal brauchst Du dabei nicht.)
Wenn man baut, soll nichts umfallen. Zeichne Deine Figuren auch so, daß die Klötze stehen bleiben.
Versuche nun, möglichst viele verschiedene Figuren zu zeichnen.



Bitte nicht umblättern !

Verrückte Geschichten

Jetzt wollen wir uns einmal ganz verrückte Fragen überlegen. Stell Dir einmal vor, es würde keinen Winter mehr geben. Was würde dann alles geschehen? Man könnte z.B. nicht mehr Schlitten fahren, die Leute könnten ihre Winterkleider wegwerfen, der Kohlenhändler müßte sich eine andere Arbeit suchen usw.

2. Nimm einmal an, Du könntest nicht nur stehen, gehen, sitzen und liegen, sondern Du könntest auch frei in der Luft schweben. Was könnte da alles geschehen ?

3. Nun stell Dir einmal vor, Wissenschaftler hätten Tabletten erfunden, die alle anderen Nahrungsmittel überflüssig machen. Die Menschen müßten nur noch täglich 3 Tabletten einnehmen und brauchten sonst nichts zu essen und zu trinken. Was könnte sich daraus alles ergeben ?

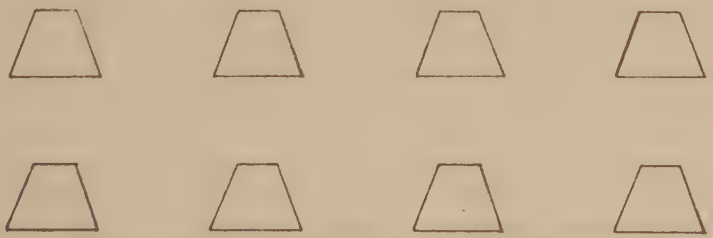
Bitte nicht umblättern !

Zeichnungen

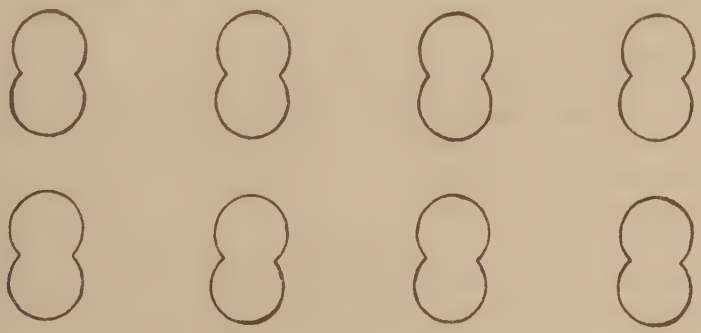
Nun kommen wir zum nächsten Spiel. Du siehst hier immer die gleichen Figuren. Male etwas dazu, damit aus den Figuren irgendwelche Dinge werden. Es kommt nicht darauf an, daß Du sehr schön zeichnest, sondern es ist wichtig, daß Dir möglichst viele verschiedene Dinge einfallen. Damit wir erkennen können, was Du gemeint hast, schreibe zu jeder Zeichnung dazu, was es sein soll.

4.

5.



6.



Bitte nicht umblättern !

Lustige Vergleiche

In diesem Spiel sollst Du herausfinden, was an zwei verschiedenen Dingen alles gleich ist.

Beispiel: Was ist bei einem Apfel und einer Orange alles gleich ?
Man kann beide essen, beide sind rund,
mit beiden kann man werfen,
beide sollten saftig sein,
beide haben eine Schale,
beide können faulen,
aus beiden kann man Köpfe basteln,
beide wachsen auf Bäumen,
beide kann man einen Abhang hinunterrollen lassen usw.

7. Nun überlege Dir, was bei einer Kerze und einem Ofen alles gleich ist.

8. Und jetzt schreib bitte auf, was bei einer Katze und einer Maus alles gleich ist.

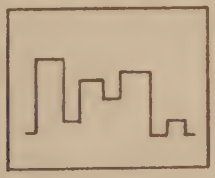
Bitte nicht umblättern !

Bilderraten

Zum Schluß haben wir noch drei Bilder, die nicht fertiggezeichnet wurden. Schau Dir die Bilder genau an und schreibe dann darunter, was das sein könnte, wenn die Zeichnung fertig wäre. Laß Deiner Phantasie ruhig freien Lauf und schreibe möglichst viele Dinge auf !

9.

10.



Bitte nicht umblättern !

11.

Was könnte das sein, wenn die Zeichnung fertig wäre ?

A7

Auswertungsanleitung

1. Allgemeine Richtlinien zur Auswertung

Die **VI,K-Aufgaben** werden nach der Anzahl der akzeptablen Antworten ausgewertet.
Da jede der 11 Aufgaben mit einer Vielzahl von Lösungen beantwortet werden kann und da die verschiedenen Lösungsmöglichkeiten weder vorhergesehen noch aufgelistet werden können, kann nur mit Hilfe von allgemeinen Angaben und Beispielen erläutert werden, was unter einer akzeptablen Antwort zu verstehen ist. In Zweifelsfällen, die bei dieser Art von Auswertung nicht völlig auszuschließen sind, sollte überlegt werden, ob die Antwort in einem erkennbaren Zusammenhang zur Aufgabe steht, wobei die kognitiven Voraussetzungen 9- bis 12jähriger Kinder zu berücksichtigen sind. Falls ein solcher Zusammenhang hergestellt werden kann, wird die Antwort gewertet, auch wenn sie vom Standpunkt Erwachsener logisch und/oder sachlich nicht ganz richtig ist. Die Erläuterungen und Beispiele zu den 11 Aufgaben werden diese Forderung verdeutlichen.

Inhaltlich übereinstimmende Antworten (Synonyme) werden nur einmal gewertet.

Wenn ein Proband denselben Gedanken mit leichten Abwandlungen mehrmals wiederholt, erhält er maximal 3 Punkte, d.h. mehr als 2 solcher Variationen desselben Einfalls werden nicht gewertet. (Beispiele für Abwandlungen der gleichen Antwort werden zu den Aufgaben, bei denen sie aufgetreten sind, gegeben.)

Treten ein Oberbegriff und ein oder mehrere zu dieser Klasse gehörende Begriffe auf - z.B. Auto, Pkw, Lkw, Ford, VW -, so werden insgesamt nur bis zu 3 Beispiele gewertet; der Oberbegriff erhält in diesem Fall keinen Punkt.
In dem hier angeführten Beispiel würden "Ford" und "VW" als zur Klasse der Pkw gehörig mit je einem Punkt gewertet. Für "Lkw" würde ebenfalls ein Punkt vergeben. "Auto" und "Pkw" würden als Oberbegriffe nicht gewertet. (Dieses Beispiel wird dadurch kompliziert, daß Lkw ebenfalls ein Oberbegriff ist und es auch Ford- und VW-Lastwagen gibt; falls jedoch der Schüler an Ford- und VW-Lastwagen gedacht haben sollte, gibt es dafür ebenfalls 2 Punkte und einen weiteren für "Pkw", so daß sich an dem Ergebnis nichts ändert.)

Bei der Auswertung mehrerer Testhefte soll unbedingt aufgabenweise vorgegangen werden, d.h. der Auswerter informiert sich über die Bewertung der Aufgabe 1 und wertet dann die Lösungen aller Schüler zu dieser Aufgabe aus. Dieser Vorgang wird bei Aufgabe 2 wiederholt, dann wird Aufgabe 3 vorgenommen etc.

2. Auswertung der 11 Aufgaben

Die vorangegangenen allgemeinen Richtlinien werden nun durch die Besprechung der Auswertung jeder einzelnen Aufgabe mit Hilfe von Beispielen präzisiert und ergänzt.

Spielklötze:

Jede Zeichnung, bei der 3 längliche Klötze irgendwie angeordnet sind, erhält einen Punkt. Die 3 Klötze müssen nicht die gleiche Größe haben wie die Beispiel-Klötze im Testheft. Sie müssen nicht rechtwinklig angeordnet sein. Sie können sich berühren, überschneiden oder mit einem Abstand gezeichnet sein. Ebenso erhalten Zeichnungen dieser Art je einen Punkt, da die Klötze auch von dieser Seite betrachtet werden können.

Verrückte Geschichten, Aufgabe 2:

Zusätzlich zu den allgemeinen Auswertungsrichtlinien ist bei der Auswertung der Aufgabe 2 folgendes zu beachten:

1. Für "schweben" oder "fliegen" ohne Zusatz gibt es keinen Punkt.
2. Für Aufzählungen, über welche Dinge man schweben oder fliegen könnte, gibt es insgesamt höchstens 3 Punkte; weitere Abwandlungen dieses Einfalls bleiben unberücksichtigt.
3. Für Aufzählungen, welche Verkehrsmittel überflüssig würden, sind wegen der Heterogenität der dabei zu berücksichtigenden Dinge maximal 4 Punkte zu vergeben. Dabei gibt es für alle fliegenden Verkehrsmittel wie Flugzeug, Hubschrauber, fliegende Untertasse u.a. insgesamt einen Punkt, für alle auf Rädern fahrende Verkehrsmittel wie Auto, Fahrrad, Eisenbahn etc. einen Punkt, für alle Wasserfahrzeuge wie Schiffe, Boote u.a. einen Punkt, ebenso für alle Möglichkeiten zur Überwindung von

Höhenunterschieden wie Gondel, Aufzug, Skilift u.a. einen Punkt. Beispiele für weitere akzeptable Antworten: "Man könnte an der Decke etwas reparieren" (hier wäre "ohne Leiter" zu ergänzen), "ich könnte in der Luft schlafen", "man könnte in die Vergangenheit (Zukunft, zu Gott, in den Himmel) fliegen".

4. Antworten, die davon ausgehen, daß nicht nur der Schüler, sondern alle Leute in der Luft schweben können, werden ebenfalls gewertet. Somit erhält z.B. die Antwort "Piloten würden arbeitslos" einen Punkt. Wenn eine Aufzählung weiterer Berufsgruppen folgt, die arbeitslos würden (Busfahrer, Taxifahrer, Lokomotivführer etc.), werden maximal 3 Punkte vergeben. Beispiele für nichtakzeptable Antworten: "Man könnte zur Schule gehen", "man könnte Schnitzel essen".

Verrückte Geschichten, Aufgabe 3:

Für Aufzählungen von nicht mehr benötigten Lebensmitteln gibt es insgesamt höchstens 3 Punkte. Das gleiche gilt für die Aufzählung von nicht mehr benötigtem Geschirr (Messer, Gabel, Löffel, Teller, Tassen, Schüsseln, Töpfe etc.).
Beispiele für akzeptable Antworten: "Man bräuchte keine Tiere mehr", "Schweine würden verenden" (offenbar, weil sie nicht mehr gefüttert werden, wenn sie als Nahrungsmittel nicht mehr benötigt werden), "die Erde würde uns überwuchern" (vermutlich, weil Felder und Wiesen nicht mehr genutzt werden; obwohl nicht ganz klar ist, ob an diese Begründung gedacht wurde, sollte diese Antwort gewertet werden, da ein Zusammenhang zur Frage leicht herstellbar ist), "man müßte Fabriken bauen, um alle Nahrungsmittel zu zerquetschen" (offenbar wurde angenommen, daß die vorhandenen Nahrungsmittel vernichtet werden müssen, um Fäulnis zu verhindern, Platz zu gewinnen o.ä.), "die Apotheker würden reich", "die Getreidespeicher der Bauern würden leer bleiben" usw.

Beispiele für nichtakzeptable Antworten:
"Die Autos könnten nicht mehr fahren", "die Menschen würden verhungern bzw. sterben" (falls keine einleuchtende und der Aufgabenstellung nicht widersprechende Begründung für den Tod gegeben wird; eine solche Begründung wäre z.B. " in den Tabletten sind keine Vitamine" o.ä.). Krankheiten und Gebrechen, deren Entstehung *nicht* begründet wird, erhalten keinen Punkt. Jede Antwort, in der eine gesundheitsschädigende Wirkung der Tabletten mit Begründung genannt wird (z.B. "wegen Vitaminmangel fallen die Haare aus", "wenn man nicht mehr kauen muß, gehen die Zähne kaputt", "man schrumpft ein, weil man nichts mehr trinkt"), wird gewertet. Werden zu *einer* Begründung mehrere gesundheitliche Konsequenzen angegeben, gibt es nur *einen* Punkt.

Zeichnungen:

Jede Zeichnung, die die Grundfigur einbezieht, erhält einen Punkt. Maximal sind also 2 x 16 und 1 x 12 Punkte für die 3 Aufgaben zu erreichen. Wenn die Zeichnung zu erkennen ist, wird sie gewertet, auch wenn keine Bezeichnung dabeisteht. Ebenso erhält sie einen Punkt, wenn sie nur mit Hilfe der Bezeichnung erkennbar ist.
Wenn gar nichts gezeichnet wurde, aber eine akzeptable Bezeichnung neben oder unter der Figur steht, gibt es einen Punkt. Eine zusätzliche Zeichnung muß aber in diesem Fall überflüssig sein; s. dazu Beispiele 1) und 6).

Wenn immer derselbe Gegenstand mit Abwandlungen dargestellt wird, werden maximal 3 Punkte vergeben; s. dazu Beispiele 2), 7) und 8). Wenn sich die Zeichnungen gar nicht unterscheiden, sondern nur verschieden benannt wurden, wird insgesamt nur 1 Punkt vergeben, s. Beispiele 3) und 8).

Wenn mehrere der 16 bzw. 12 Grundfiguren in einer Zeichnung verwendet werden und in dieser Zeichnung eine echte Funktion haben, werden alle verwendeten Figuren mit je einem Punkt gewertet; s. dazu Beispiel 4) bei Aufgabe 4.

Wenn aus den Figuren dieser Aufgaben - häufig bei Aufgabe 4, gelegentlich auch bei Aufgabe 5 - Bilderrahmen gemacht werden, in die etwas hineingemalt wird, so erhält nur *ein* solches "Bild" oder "Gemälde" einen Punkt; alle weiteren Abwandlungen dieses Einfalls werden nicht gewertet.

Für dekorative Muster, die aus den Grundfiguren durch Hinzufügungen gezeichnet wurden, gibt es keinen Punkt.

Beispiele zu den Aufgaben 4 - 6:

Aufgabe 4

1) Spiegel = 1 Punkt
   Teller = 1 Punkt
   Auge = 0 Punkte, wenn nichts gezeichnet wurde.

2) Teller mit Suppe
   Teller mit Brei
   Teller mit Schnitzel
   Teller mit Gemüse ... = 3 Punkte, auch bei 16 verschiedenen Gerichten

3) Verschieden benannte, aber gleich gezeichnete Tiere werden nur einmal gewertet. Wenn die Figur der Aufgabe 4 immer der Rumpf ist, an den ein Kopf, 4 Beine und ein Schwanz angehängt werden, gibt es dafür nur einen Punkt, auch wenn "Hund", "Katze", "Löwe" und "Pferd" dazugeschrieben werden. Wenn jedoch ein Fisch, ein Maikäfer, ein Tausendfüßler und ein Schwein gezeichnet werden, gibt es 4 Punkte.

4) "Lampe" = 2 Punkte

Aufgabe 5

5) Die Figuren der Aufgabe 5 können auch auf den Kopf gestellt in den Zeichnungen verwendet werden. Beispiel: "Blumentopf" = 1 Punkt.

6) "Startblock" = 1 Punkt

7) Für "Eimer", "Schüssel", "Blumentopf", "Tasse", "Becher" und "Topf" gibt es maximal 3 Punkte, wenn es sich um leicht abgewandelte Zeichnungen handelt. (Wenn nichts oder immer dasselbe gezeichnet wurde, aber verschiedene Benennungen dazugeschrieben wurden, wird insgesamt nur 1 Punkt vergeben.)

Aufgabe 6

8) Wenn aus der Figur der Aufgabe 6 auf immer gleiche oder sehr ähnliche Art Menschen gezeichnet werden - Vater, Mutter, Kind -, wird nur ein Punkt vergeben. Wenn allerdings bei einer dieser Figuren Kopf, Arme und Beine - mit oder ohne Kleider - dazugezeichnet werden und aus einer anderen ein "Wickelkind" gemacht wird, werden beide Zeichnungen gewertet, da die Ausgangsfigur in unterschiedlicher Weise für die Zeichnung verwendet wird.

Lustige Vergleiche

Zusätzlich zu den allgemeinen Auswertungsrichtlinien ist folgendes zu beachten: Im Zweifelsfall sollte sich der Auswerter nicht fragen: "Trifft die Antwort immer für beide Objekte bzw. für beide Tiere zu?", sondern: *"Kann* das für beide zutreffen - und sei es nur in Einzelfällen?" Die Beispiele zu den beiden Aufgaben werden diese Forderung verdeutlichen.

Mehrere inhaltlich übereinstimmende Begriffe - wie z.B. Feuer und Flamme, brennen und glühen - erhalten zusammen nur einen Punkt.

Akzeptable Antworten zu Aufgabe 7: "Beides geht nach einer Zeit kaputt." (Zwar gehen eine Kerze nur durch Verbrauch, also sehr rasch, und ein Ofen nur durch Verschleiß, also nach einer vergleichsweise sehr langen Zeitspanne, kaputt, aber die Antwort ist natürlich richtig und wird gewertet.)
"Beides qualmt." (Das muß nicht notwendigerweise, kann jedoch der Fall sein, die Antwort wird also gewertet.)

Akzeptable Antworten zu Aufgabe 8: "Beide haben ein Fell, 4 Beine, Augen, Ohren, Magen" etc; solche anatomischen Analogien erhalten maximal 3 Punkte. "Haustiere" (hier soll nicht der richtige Gebrauch des Begriffs "Haustiere" geprüft werden, sondern ob die Pbn Gemeinsamkeiten finden können; da beide Tiere in Häusern leben oder jedenfalls leben *können*, erhält diese Antwort einen Punkt.)

"Mit beiden kann man spielen." (Diese Antwort scheint zunächst falsch zu sein, da Mäuse im allgemeinen weglaufen, wenn Kinder versuchen, mit ihnen zu spielen; sie wird jedoch gewertet, da es durchaus möglich ist, Mäuse durch Futter anzulocken und so zu dressieren, daß ein "Spiel" denkbar ist.)
"Beide haben das Katz-und-Maus-Spiel erfunden."(Sicher bedarf es keiner Begründung, warum diese "falsche" Antwort in einem Kreativitätstest gewertet wird!)

Bilderraten

Zusätzlich zu den allgemeinen Auswertungsrichtlinien ist bei diesen Aufgaben besonders zu berücksichtigen, daß es für einen Oberbegriff und dazugehörige Beispiele oder Spezialfälle insgesamt nur einen Punkt gibt. Außerdem wird daran erinnert, daß synonyme Begriffe nur *einmal* gewertet werden.
Pars-pro-toto-Deutungen werden gewertet, auch wenn die vollständige Zeichnung *nicht* in den vorgegebenen Rahmen passen würde. (Beispiel: "Baum" als Lösung von Aufgabe 9, obwohl nur die Krone zu sehen ist und der vollständige Baum nicht in den vorgegebenen Rahmen gezeichnet werden könnte.)
Die Antworten "ein Bild", "ein Kunstwerk", "ein Muster" bzw. "es kann ein Bild sein" u.ä. dürfen nur einmal bei den Aufgaben 9, 10 und 11 gewertet werden, falls sie mehrmals gegeben werden.

Für die Auswertung dieser - wie auch der übrigen - Aufgaben benötigt der Auswerter selbst etwas Kreativität. In Zweifelsfällen, d.h. wenn nicht sicher ist, ob der Pb etwas "Richtiges" ge-

meint hat, sollte ein Punkt immer dann gegeben werden, wenn er etwas Richtiges gemeint haben *könnte*.

Akzeptable Antworten zu Aufgabe 9:
"Strauch" oder "Busch" ("Strauch", ..., "Busch" erhalten als inhaltlich übereinstimmende Begriffe zusammen nur einen Punkt.) "Reifen", ..., "Rad" werden getrennt gewertet, da mit Rad auch Zahnrad u.a. gemeint sein kann.
"Wolke", ..., "Staubwolke" erhalten je einen Punkt. (Wolke ist eine überaus häufige, Staubwolke eine sehr seltene Antwort.)
"Kopf", ..., "Haare" erhalten zusammen nur einen Punkt, weil die Deutung in beiden Fällen von den Haaren ausgeht.
"Ball" sowie alle Gegenstände, die zwar rund, aber ganz glatt sind, werden nicht gewertet.

Akzeptable Antworten zu Aufgabe 10:
"Burg", "Burgruine", "Festung", "Ruine", "Schloß" erhalten insgesamt einen Punkt, auch wenn ein Pb mehrere dieser Antworten gegeben hat, da immer Burg(-ruine) gemeint ist. "Großstadt", "Häuser", "Hochhäuser", "Stadt", "Wohnblöcke", "Wolkenkratzer" werden insgesamt nur einmal gewertet, auch wenn mehrere dieser Begriffe aufgeschrieben wurden, da die Deutung immer von den Silhouetten verschieden hoher Häuser ausgeht.
"Fabrik", "Fabrikschornsteine", "Schornsteine" werden ebenfalls nur einmal gewertet.
"Entwurf", "technische Zeichnung", "Zeichenform" u.ä. erhalten insgesamt nur einen Punkt.
Garagen = 1 Punkt (offenbar wurde die Zeichnung als eine Sicht von oben auf Garagen für verschieden lange Fahrzeuge - Lastzüge, Pkw - gedeutet).
Geschnittene Hecke = 1 Punkt
"Eckige Bäume" = 1 Punkt (vermutlich wurde an Bäume mit geschnittenen Kronen, wie man sie gelegentlich noch auf alten Friedhöfen findet, gedacht - oder das Wort "Hecke" fiel dem Pbn gerade nicht ein.) Wenn beide Antworten zusammen auftreten, werden sie nur einmal gewertet.

Akzeptable Antworten zu Aufgabe 11:
"Berg", "Zugspitze" sowie andere Namen von Bergen erhalten insgesamt einen Punkt.
"Brief", "Briefumschlag", "Briefkuvert", "Kuverte" sind Deutungen der Zeichnung *mit* Rahmen und werden gewertet. Gibt ein Pb mehrere dieser Antworten, erhält er nur einen Punkt.
Bumerang = 1 Punkt
Dreieck = 1 Punkt
ein K = 1 Punkt (auf der Seite liegendes K!)
Pfeilspitze = 1 Punkt (schreibt der Pb noch weitere "Spitzen", z.B. Speer- und Schwertspitze, erhält er insgesamt nur einen Punkt)
Richtungspfeil = 1 Punkt
Straße = 1 Punkt (Straße, die vom Betrachter wegführt und in der Ferne verschwindet, also aus perspektivischer Sicht)
Zelt = 1 Punkt

Anhang B

Anhang B 1

Ergebnisse der Untersuchungen zur Auswertungsobjektivität

Die Frage, ob die zu den 11 Aufgaben ausgearbeiteten und mehrfach modifizierten bzw. ergänzten Auswertungsrichtlinien eine weitgehend objektive Auswertung des Tests ermöglichen, wurde dadurch überprüft, daß die Testhefte von 1o Klassen, die nach Zufall aus der Gesamtstichprobe von insgesamt 57 Klassen ausgewählt worden waren, zwei Mal ausgewertet wurden und zwar von jeweils verschiedenen Auswertern.
Tabelle I zeigt die Produkt-Moment-Korrelationen zwischen den Ergebnissen der beiden Auswertungen.

Tabelle I: Produkt-Moment-Korrelationen für die Ergebnisse zweier Auswertungen derselben Testhefte. (N = 3oo)

Aufgabe	r (Anzahl)	r (Originalität)
1	.94	-
2	.85	.76
3	.85	.79
4	.83	.82
5	.95	.89
6	.93	.9o
7	.95	.93
8	.73	.85
9	.92	.84
1o	.85	.73
11	.79	.78
Gesamtwert	.93	.9o

Für einen Test mit freier Aufgabenbeantwortung und einer Vielzahl möglicher Lösungen sind die Koeffizienten sehr hoch.
In einer weiteren Prüfung der Auswertungsobjektivität wurden die Testhefte von 2 Schulklassen von 3 unabhängig arbeitenden Auswertern nach der Gesamtzahl der akzeptablen Lösungen ausgewertet. Die Korre-

lation zwischen den Ergebnissen je zweier Auswerter erhärten die günstigen Ergebnisse der Tabelle I :

$r_{1/2} = .96; \quad r_{1/3} = .95; \quad r_{2/3} = .98;$

Da sich sehr hohe Korrelationen auch dann ergeben, wenn einer der Auswerter konsistent strenger oder großzügiger auswertet, wurden auch die arithmetischen Mittel der Gesamt-Anzahlwerte auf signifikante Unterschiede überprüft. Tabelle II zeigt die Ergebnisse.[1)]

Tabelle II: Mittelwerte der Produktivitätsscores für 2 Klassen, ausgewertet durch 3 unabhängige Auswerter, sowie die t-Werte und Zufallswahrscheinlichkeiten p der Unterschiede (T - 58)

Mittelwerte je zweier Auswerter	t-Wert für abhängige Stichproben	p
$\bar{X}_1 = 66,98, \bar{X}_2 = 67,28$	0.48	0.634
$\bar{X}_1 = 66,98, \bar{X}_3 = 64,84$	3.12	0.003 ss*
$\bar{X}_2 = 67,28, \bar{X}_3 = 64,84$	5.07	0.001 sss*

*) Sehr signifikante t-Werte ($p \leq .01$) sind durch ss, t-Werte mit $p \leq .001$ sind durch sss gekennzeichnet.

Während die arithmetischen Mittel der Anzahlwerte der Auswerter 1 und 2 sich kaum unterscheiden, liegt der Mittelwert des 3. Auswerters hochsignifikant unter denjenigen der beiden anderen Auswerter. Hier sind weitere Untersuchungen erforderlich, um zu klären, ob solche systematischen Unterschiede bei genauer Befolgung der Auswertungsrichtlinien öfter auftreten oder ob der 3. Auswerter dieser Untersuchung eine Ausnahme darstellt.

---

1) Programm DIFT des Deutschen Rechenzentrums; Autor: Dr. F.Gebhardt, DRZ Darmstadt

Anhang B 2

Tab. III: Überblick über die Zusammenhänge zwischen Kreativität bzw. "giftedness" und den CPQ-Faktoren in den Untersuchungen von Porter (1964), Werner (1966) und Werner und Bachtold (1969). Die zu diesen Arbeiten mitgeteilten Zahlen sind die Wahrscheinlichkeiten für den Unterschied zwischen dem jeweiligen Faktor-Mittelwert der hochbegabten Kinder und demjenigen der Normgruppe

Faktoren

Autor(en)	A	B	C	D	E	F	G	H	I	J	N	O	Q_3	Q_4
Porter 1964														
31 Jungen		+.01					+.01*		-.01*		-**			
35 Mädchen	+.01	+.01			-.01		+.01	+.05			-		+.01	
Gesamtgruppe	+.01	+.01		-.05	-.02		+.01	+.02			-	-.05	+.05	
Werner 1966														
2o Jungen	-.001	+.001	+.001		+.1o		-.1o	+.001				-.01		
23 Mädchen	+.1o	+.001			-.01		+.01				-.1o			
Werner und Bachtold 1969														
113 Jungen	+.001	+.001	+.001		+.01		+.001	+.001	+.01	+.01	-.001	-.001	+.001	-.05
98 Mädchen	+.001	+.001	+.001				+.05	+.001			-.001	-.001	+. 01	
57 Jungen	-.1o	+.001			+.001					+.001	-**	-.02	+.001	
56 Mädchen		+.001				+.001	-.01			+.1o	-			
Cattel und Butcher 1968***	-.3A	+.6B	+.3C		+.3E	-.2F	+.1G	+.1H	-.1I			-.10		

*)+bedeutet, daß der Mittelwert der hochbegabten Schüler höher, − bedeutet, daß er niedriger ist als derjenige der Normgruppe.

**)Die Stichprobe von Porter und die zweite Stichprobe von Werner und Bachtold wurden nicht mit dem CPQ, sondern mit dem HSPQ (High School Personality Questionnaire) untersucht, der statt des im CPQ gemessenen Faktors N den Faktor Q_2 erfaßt.

***)Um den Überblick möglichst zu vervollständigen, werden zusätzlich Vorzeichen und Regressionskoeffizienten derjenigen CPQ-Faktoren mitgeteilt, die Cattell und Butcher in ihrer Spezifikationsgleichung für die Vorhersage der Kreativität von Schülern berücksichtigen.

Anhang B 3

Tabelle IV: Arithmetisches Mittel ($\bar{X}$), Standardabweichung (δ) und Signifikanz der Abweichung von der Normalverteilung (AbwNV) für die Anzahl- und Originalitätswerte der Items und für die Anzahl- und Originalitäts-Gesamtwerte der Versuchsform VI,K*

	Gesamtschulen N = 303	Realschulen N = 408	Gymnasien N = 296	Hauptschulen N = 94	Gesamtstichprobe N = 1671
Aufgabe 1 Anzahl	$\bar{X}$ = 6,5- δ = 3,9- AbwNV = ss**	$\bar{X}$ = 6,5- δ = 4,17 AbwNV = ss	$\bar{X}$ = 12,00 δ = 6,43 AbwNV = ss	$\bar{X}$ = 7,41 δ = 4,30 AbwNV = ss	$\bar{X}$ = 7,40 δ = 4,28 AbwNV = ss
Aufgabe 2 Anzahl	$\bar{X}$ = 4,87 δ = 2,24 AbwNV = ss	$\bar{X}$ = 5,07 δ = 2,14 AbwNV = s**	$\bar{X}$ = 5,85 δ = 2,44 AbwNV = ss	$\bar{X}$ = 5,73 δ = 2,72 AbwNV = ns**	$\bar{X}$ = 4,88 δ = 2,36 AbwNV = ss
Originalität	$\bar{X}$ = 6,09 δ = 3,74 AbwNV = ss	$\bar{X}$ = 7,49 δ = 3,49 AbwNV = ns	$\bar{X}$ = 8,95 δ = 4,24 AbwNV = ss	$\bar{X}$ = 7,34 δ = 4,45 AbwNV = ns	$\bar{X}$ = 7,01 δ = 3,96 AbwNV = ss
Aufgabe 3 Anzahl	$\bar{X}$ = 4,06 δ = 2,02 AbwNV = ss	$\bar{X}$ = 4,69 δ = 2,27 AbwNV = ss	$\bar{X}$ = 5,19 δ = 2,65 AbwNV = ss	$\bar{X}$ = 4,19 δ = 2,25 AbwNV = ss	$\bar{X}$ = 4,42 δ = 2,26 AbwNV = ss
Originalität	$\bar{X}$ = 4,58 δ = 3,04 AbwNV = ss	$\bar{X}$ = 5,49 δ = 3,05 AbwNV = ss	$\bar{X}$ = 5,64 δ = 3,87 AbwNV = ss	$\bar{X}$ = 4,20 δ = 2,92 AbwNV = ss	$\bar{X}$ = 4,97 δ = 3,24 AbwNV = ss
Aufgabe 4 Anzahl	$\bar{X}$ = 5,26 δ = 2,17 AbwNV = ss	$\bar{X}$ = 6,00 δ = 1,90 AbwNV = ss	$\bar{X}$ = 6,85 δ = 1,64 AbwNV = ss	$\bar{X}$ = 5,77 δ = 1,86 AbwNV = ns	$\bar{X}$ = 5,76 δ = 2,08 AbwNV = ss
Originalität	$\bar{X}$ = 6,74 δ = 3,27 AbwNV = ss	$\bar{X}$ = 7,76 δ = 3,07 AbwNV = ss	$\bar{X}$ = 8,81 δ = 3,21 AbwNV = ns	$\bar{X}$ = 7,04 δ = 2,93 AbwNV = ns	$\bar{X}$ = 7,38 δ = 3,29 AbwNV = ss
Aufgabe 5 Anzahl	$\bar{X}$ = 4,46 δ = 2,16 AbwNV = ss	$\bar{X}$ = 5,11 δ = 1,97 AbwNV = ss	$\bar{X}$ = 6,27 δ = 1,76 AbwNV = ss	$\bar{X}$ = 5,60 δ = 1,95 AbwNV = s	$\bar{X}$ = 5,01 δ = 2,14 AbwNV = ss
Originalität	$\bar{X}$ = 4,86 δ = 3,36 AbwNV = ss	$\bar{X}$ = 5,52 δ = 3,21 AbwNV = ss	$\bar{X}$ = 7,44 δ = 3,39 AbwNV = ss	$\bar{X}$ = 6,77 δ = 3,48 AbwNV = ns	$\bar{X}$ = 5,59 δ = 3,48 AbwNV = ss
Aufgabe 6 Anzahl	$\bar{X}$ = 4,10 δ = 1,20 AbwNV = ss	$\bar{X}$ = 4,43 δ = 1,95 AbwNV = ss	$\bar{X}$ = 5,75 δ = 1,99 AbwNV = ss	$\bar{X}$ = 4,65 = 2,13 AbwNV = ns	$\bar{X}$ = 4,50 δ = 2,08 AbwNV = ss
Originalität	$\bar{X}$ = 5,49 δ = 3,35 AbwNV = ss	$\bar{X}$ = 6,49 δ = 3,40 AbwNV = ss	$\bar{X}$ = 8,51 δ = 3,73 AbwNV = ss	$\bar{X}$ = 6,49 δ = 3,61 AbwNV = ns	$\bar{X}$ = 6,33 δ = 3,62 AbwNV = ss
Aufgabe 7 Anzahl	$\bar{X}$ = 4,33 δ = 1,96 AbwNV = ss	$\bar{X}$ = 5,36 δ = 2,10 AbwNV = s	$\bar{X}$ = 6,16 δ = 2,84 AbwNV = ss	$\bar{X}$ = 4,64 δ = 1,96 AbwNV = s	$\bar{X}$ = 4,92 δ = 2,28 AbwNV = ss
Originalität	$\bar{X}$ = 3,53 δ = 2,79 AbwNV = ss	$\bar{X}$ = 4,63 δ = 3,39 AbwNV = ss	$\bar{X}$ = 6,06 δ = 4,49 AbwNV = ss	$\bar{X}$ = 4,11 δ = 2,86 AbwNV = ns	$\bar{X}$ = 4,28 δ = 3,43 AbwNV = ss
Aufgabe 8 Anzahl	$\bar{X}$ = 5,48 δ = 2,50 AbwNV = ss	$\bar{X}$ = 7,76 δ = 3,40 AbwNV = ss	$\bar{X}$ = 8,44 δ = 3,69 AbwNV = ss	$\bar{X}$ = 7,06 δ = 2,51 AbwNV = ns	$\bar{X}$ = 6,65 δ = 3,23 AbwNV = ss
Originalität	$\bar{X}$ = 5,28 δ = 3,55 AbwNV = ss	$\bar{X}$ = 6,97 δ = 4,31 AbwNV = ss	$\bar{X}$ = 8,48 δ = 5,46 AbwNV = ss	$\bar{X}$ = 6,51 δ = 3,68 AbwNV = ns	$\bar{X}$ = 6,33 δ = 4,31 AbwNV = ss
Aufgabe 9 Anzahl	$\bar{X}$ = 6,61 δ = 3,13 AbwNV = ss	$\bar{X}$ = 8,29 δ = 4,02 AbwNV = ss	$\bar{X}$ = 8,82 δ = 4,64 AbwNV = ss	$\bar{X}$ = 6,13 δ = 2,79 AbwNV = ss	$\bar{X}$ = 7,39 δ = 3,78 AbwNV = ss
Originalität	$\bar{X}$ = 6,70 δ = 4,69 AbwNV = ss	$\bar{X}$ = 8,35 δ = 6,22 AbwNV = ss	$\bar{X}$ =10,10 δ = 7,73 AbwNV = ss	$\bar{X}$ = 6,03 δ = 3,94 AbwNV = s	$\bar{X}$ = 7,67 δ = 5,85 AbwNV = ss

Tabelle IV (Fortsetzung)

	Gesamt-schulen N = 373	Real-schulen N = 408	Gymnasien N = 296	Haupt-schulen N = 94	Gesamt-stichprobe N = 1671
Aufgabe 1o Anzahl	$\bar{X}$ = 4,87 ó = 2,43 AbwNV = ss	$\bar{X}$ = 5,49 ó = 2,39 AbwNV = ss	$\bar{X}$ = 5,62 ó = 2,96 AbwNV = ss	$\bar{X}$ = 4,64 ó = 2,3o AbwNV = ss	$\bar{X}$ = 5,14 ó = 2,54 AbwNV = ss
Originalität	$\bar{X}$ = 3,53 ó = 3,55 AbwNV = ss	$\bar{X}$ = 4,4o ó = 3,78 AbwNV = ss	$\bar{X}$ = 4,73 ó = 4,57 AbwNV = ss	$\bar{X}$ = 2,91 ó = 2,95 AbwNV = ss	$\bar{X}$ = 3,92 ó = 3,81 AbwNV = ss
Aufgabe 11 Anzahl	$\bar{X}$ = 5,89 ó = 3,o5 AbwNV = ss	$\bar{X}$ = 7,68 ó = 3,65 AbwNV = ss	$\bar{X}$ = 7,21 ó = 4,o1 AbwNV = ss	$\bar{X}$ = 5,19 ó = 2,32 AbwNV = ns	$\bar{X}$ = 6,52 ó = 3,46 AbwNV = ss
Originalität	$\bar{X}$ = 6,o9 ó = 4,88 AbwNV = ss	$\bar{X}$ = 8,23 ó = 5,91 AbwNV = ss	$\bar{X}$ = 8,34 ó = 6,66 AbwNV = ss	$\bar{X}$ = 5,33 ó = 3,45 AbwNV = s	$\bar{X}$ = 6,97 ó = 5,54 AbwNV = ss
Anzahl-Gesamtwert	$\bar{X}$ =56,19 ó =17,52 AbwNV•••	$\bar{X}$ =66,42 ó nicht berechnet	$\bar{X}$ =78,16 ó nicht berechnet	$\bar{X}$ =61,o1 ó nicht berechnet	$\bar{X}$ =62,6o ó =2o,38
Originalitäts-Gesamtwert	$\bar{X}$ =52,88 ó =21,8o	$\bar{X}$ =65,33 ó nicht berechnet	$\bar{X}$ =77,o6 ó nicht berechnet	$\bar{X}$ =56,73 ó nicht berechnet	$\bar{X}$ =6o,42 ó =25,35
Gesamtwert aus Anzahl- und Originalitätswerten	$\bar{X}$ = 1o9,o7 ó = 38,44	$\bar{X}$ = 131,75 ó nicht berechnet	$\bar{X}$ = 155,22 ó nicht berechnet	$\bar{X}$ = 117,74 ó nicht berechnet	$\bar{X}$ = 123,o2 ó = 44,87

*) Programm NRMP des Deutschen Rechenzentrums; Autor: Dr. F. Gebhardt, DRZ Darmstadt
**) s bedeutet $p \leq .o2$, ss bedeutet $p \leq .oo5$, ns heißt $p > .o2$.
***) Für die Gesamtwerte wurde keine Normalitätsprüfung vorgenommen.

Tabelle IV zeigt, daß sich die schultypspezifischen Mittelwerte stark unterscheiden. Die Gymnasien erreichen fast ausnahmslos die höchsten Ergebnisse, meist gefolgt von den Realschulen; überraschend ist das vergleichsweise gute Abschneiden der Hauptschüler, die die Schüler der Gesamtschulen in fast allen und die Realschüler in einigen Aufgaben übertreffen.
Aus diesem Ergebnis sollten allerdings wegen der zweifelhaften Repräsentativität der Hauptschulstichprobe - die Beteiligung an der Untersuchung erfolgte auf freiwilliger Basis - keine Schlußfolgerungen abgeleitet werden.
Die Anzahl der Originalitätspunkte scheint stark von dem jeweiligen Item abzuhängen; z.B. fallen bei der 2. Aufgabe die Originalitäts-Mittelwerte in allen Schultypen erheblich höher aus als die Anzahl-Mittelwerte, bei Aufgabe 7 ist es umgekehrt. Die Streuung der Originalitätswerte ist bei allen Aufgaben und in allen Schultypen wesentlich größer als diejenige der Produktivitätswerte.
Die meisten Verteilungen weichen signifikant von der Normalverteilung ab. Der Produktivitäts-Gesamtwert zeigt im Gegensatz zum Originalitäts-Gesamtwert keine signifikante Abweichung von der Normalverteilung.[1)]

Eine Inspektion der Häufigkeiten der verschiedenen Anzahl- und Originalitätswerte[2)] zeigt, daß die meisten Verteilungen rechtsschief bzw. positiv schief sind. Nur die Produktivitätswerte der Aufgabe 4 zeigen für alle Schultypen außer der Hauptschule ausgeprägt linksschiefe (negativ schiefe) Verteilungen, was bedeutet, daß ein Teil der Kinder mehr als die vorgegebenen 8 Figuren bearbeiten könnte. Bei den Gymnasiasten trifft dies auch für die Aufgaben 5 und 6 zu.[3)] Für diese 3 zeichnerisch zu lösenden Aufgaben ist also die Testdecke zu niedrig.
Abgesehen von diesen Aufgaben fielen der Mehrzahl der Kinder relativ wenige und überwiegend die gleichen Antworten ein.

---

1) Da Lienert (1967[2]) sowohl für Schiefe als auch für Exzeß empfiehlt, die Hypothese einer normalen Rohwertverteilung erst auf der 1 %igen Sicherheitsstufe zu verwerfen, wurden nach dem von ihm angegebenen Verfahren t-Werte für Schiefe und Exzeß berechnet und konservativ getestet, was zu einigen Abweichungen von den Ergebnissen des NRMP-Programms führte.

2) Programm HAUF des Deutschen Rechenzentrums; Autoren: P. Schnell und L. Pöcker, DRZ Darmstadt

3) Überprüfung mit dem t-Test für die Schiefe der Verteilung, s. dazu Lienert 1967[2], S. 172. Die Linksschiefe der Verteilung ist bei den Hauptschülern mit einer Irrtumswahrscheinlichkeit von 5 Prozent nicht ausreichend gesichert (vgl. Anmerkung 1) oben!), bei den anderen Schultypen ist sie mit einer Irrtumswahrscheinlichkeit von 1 Prozent signifikant.

Lebenslauf

Ursula M. Mainberger
geboren am 2o. Juni 1936 in Karlsruhe
Vater: Kurt Simon, Fabrikant
Mutter: Elisabeth Simon, geborene Dörler

1942 - 1947	Besuch der Volksschule in Wilgartswiesen/Pfalz
1947	Übertritt in das Neusprachliche Gymnasium in Landau/ Pfalz
1955	Reifeprüfung
1955 - 1957	Studium der Psychologie an den Universitäten Würzburg und Heidelberg. Abbruch des Studiums aus familiären Gründen (Tod der Mutter)
1963	Wiederaufnahme des Studiums an der Unversität des Saarlandes
1971	Diplom-Hauptprüfung in Psychologie an der Universität des Saarlandes
1971 - 1975	Wissenschaftliche Assistentin,
1975 - 1976	Wissenschaftliche Mitarbeiterin an der Erziehungswissenschaftlichen Hochschule Rheinland-Pfalz, Abteilung Landau. Mitarbeit im Zentrum für empirische pädagogische Forschung (Leitung: Prof. Dr. K. Ingenkamp)
Seit 1976	Studienberaterin an der Universität Düsseldorf
Seit 1958	verheiratet mit Diplom-Holzwirt Dr. Robert Mainberger

Druck: D. Gräbner, 8602 Altendorf – Telefon: 09545 / 8791